CHAPTER 2

LITERATURE REVIEW

2.1 REPORTED WORKS ON DIMENSIONALITY REDUCTION FOR HUMAN FACE RECOGNITION

[12] presented a system for person-independent hand posture recognition against complex backgrounds cues. It has a specific color i.e. the color of skin and it moves in the image with a characteristic speed. There are many more high-level cues, such as the form, texture or specific trajectory which characterize a moving hand. However, in order to be fast and robust, it sticks with the simplest approach that still does the job.

[18] presented component-based face recognition with 3D morphable models. When the eigenfaces have been created, each image will be represented as a vector of weights. The system is ready to accept incoming queries. The weight of the incoming unknown image is found and then compared to the weights of those already in the system. If the input image's weight is over a given threshold it is considered to be unknown the identification of the input image is done by finding the image in the database whose weights are the closest to the weights of the input image.

[22] proposed recognizing song-based blink patterns: applications for restricted and universal access. It introduced an enrollment process for an identity verification system based on blink recognition that helps increase the viability of blinking as a biometric indicator. With increased verification accuracy, blinking could be combined with face recognition to allow non-obtrusive, hands-free biometric-based identification systems to be created.

[47] presented object class recognition and localization using sparse features with limited receptive fields. The original HMAX model is designed to analyze 2D images. This model has been extended to recognize actions in video data. In particular, the Gabor filters in S1 layer of the HMAX model have been replaced with some gradient and space-time modules to capture motion information.

[62] proposed semi-supervised bi-directional dimensionality reduction for face recognition. It is based on semi-supervised learning, domain knowledge in the form of pair wise constraints besides abundant unlabeled examples is available, which specifies whether a pair of instances belong to the same class or not. Compared to the semi-supervised dimensionality reduction (SSDR), it cannot preserve the intrinsic structure of the unlabeled data as well as both the must-link and cannot-link constraints defined.

[48] proposed face recognition using contour matching, A face recognition systems developed using contour matching technique is efficient due to the fact that the exact structure of the face can be extracted by “contours”. Areas of constant grey level are enclosed using contour lines and the whole face can be treated as a contour map. The strength of this method is that the storage requirements are less because for matching purpose whole face is not needed, only extracted contours are used.

[19] proposed how should an autonomous vehicle overtake a slower moving vehicle: Design and analysis of an optimal trajectory. In the intelligent transportation systems, the extraction and tracking of objects of interest are the necessary prerequisites for the development of intelligent autonomous vehicles or mobile robots. In the case of forward collision avoidance, visual moving object tracking helps to distinguish the potential collision threats in terms of their relevance to the intended path of the vehicle.

[49] presented efficiency improvement for unconstrained face recognition by weightening probability values of modular PCA and wavelet PCA. PRPCA uses a custom designed initialization algorithm to appropriate set of initial projection vectors for the PCA. The PCs are prioritized by the projection vectors. An example of such an algorithm is Automatic Target Generation Process (ATGP) Modular PCA is a modified form of PCA. The new version divides the image into equal size sub images on which PCA is applied resulting in the formation of individual eigenfaces for each sub image.

[60] presented LDA based face recognition by using Hidden Markov Model in current trends. Hidden Markov model (HMM) is a promising method that works well for images with variations in lighting, facial expression and orientation. The face recognition draws attention as a complex task due to noticeable changes produced on appearance by illumination, facial expression, size, orientation and other external factors.

[30] proposed similarity-based online feature selection in content-based image retrieval. Content-based image retrieval (CBIR) has gap between high-level semantic concepts and low-level visual features hinders further performance improvement. The problem of online feature selection is critical to really bridge this gap. It investigates online feature selection in the relevance feedback learning process to improve the retrieval performance of the region-based image retrieval system.

[15] presented learning in content-based image retrieval context-based image retrieval (CBIR) has been largely explored in the last decade. In the CBIR context, an image is represented by a set of low-level visual features, which have no direct correlation with high-level semantic concepts and the gap between high-level concepts and low-level features is the major difficulty that hinders further development of CBIR systems.

[23] presented an effective and efficient region-based image retrieval framework. The relevance feedback mechanism is an iterative learning process, which is generally treated as online supervised learning. During each iteration, the user labels some images to be “relevant” or “irrelevant” to his query concept and the system uses the labeled images as training samples to successively refine the learning mechanism and gives better retrieval results in the next iteration. The RBIR approaches segment images into several regions and retrieve images with low-level features extracted based on these regions.

[41] presented effective lip localization and tracking for achieving multimodal speech recognition. Effective fusion of acoustic and visual modalities in speech recognition has been an important issue in Human Computer Interfaces (HCI), warranting further improvements in intelligibility and robustness. Speaker lip motion stands out as the most linguistically relevant visual feature for speech recognition. It has a new hybrid approach to improve lip localization and tracking aimed at improving speech recognition in noisy environments.

[38] presented human face recognition that do not operate directly on the reflectance of light from a face (i.e. the pixel intensity values). Instead, the primate visual system extracts visual features that are in turn processed holistically. It inspired machine object recognition systems, Gabor filters have been used to simulate these human visual features. Similar texture features such as local binary patterns and histograms of orientation gradients serve the same function, but they are more in tune with the discrete nature of digital images.

[45] proposed a visual front-end for a continuous pose-invariant lip reading system having an audio-visual automatic speech recognition (AVASR) system which can recognize what a speaker’s says regardless of head position (i.e. left profile, front, right profile etc.), would be most useful as it enables this

technology to be used in a host of realistic applications such as mobile phone and in vehicle speech recognition. A major hurdle in achieving this goal is in developing a visual front-end which can effectively locate and track a user's face and facial features from a single camera.

[39] proposed integrating audio visual data for human action detection. This approach is top-down for determine and extract action scenes in video by analyzing both audio and video data. It modeled the hierarchy and shared structures of human behaviors and a framework of the Hidden Markov model based application for the problem of activity recognition. It measures human action based information from video with the method deals with both visual and auditory information that captures both spatial and temporal characteristics.

[37] presented artificial visual system algorithm applied on face recognition and tracking. The development of an artificial visual system consists of several algorithms based on the biological human eye sight, particularly about the retina. It involves algorithms such as log polar transformation, Gaussian filters, Gabor Wavelets, etc. Overall, the system is implemented into recognition and able to perform some tracking abilities.

2.2 REPORTED WORKS ON HUMAN FACE INTERACTION

[1] proposed human-computer interaction using Eye-Gaze input. It uses as a component of a user interface, the eye movement data must be obtained in real time and used in some way that has an immediate effect on the dialogue. This situation has been studied most often for disabled (quadriplegic) users, who can use only their eyes for input. Because other user-computer communication modes are unavailable, the resulting interfaces are rather slow and tricky to use for non-disabled people, but, of course, a tremendous boon to their intended users.

[5] presented robot control using monocular pose estimation. Vision-based robot control can be classified, depending on the error used to compute the control law into four groups: position-based, image-based, hybrid and motion-based control systems. In a position-based control system, the error is computed in the 3D Cartesian space (for this reason, this approach can be called 3D visual servoing). In an image-based control system, the error is computed in the 2D image space (for this reason, this approach can be called 2D visual servoing).

[9] proposed (Some) computer vision based interfaces for interactive art and entertainment installations. This work purpose is to provide the reader with information on some of the tools available for computer vision based body tracking, how they can be selected and applied to achieve the desired artistic goal and their limitations. What these computer vision interfaces have in common is low cost and ease of implementation, as it require means which are commonly available to most individuals/institutions, such as computers and small cameras.

[61] proposed advanced recognition techniques for human computer interaction. Hand gestures are an important modality for human computer interaction (HCI) that compared to many existing interfaces, hand gestures have the advantages of being easy to use, natural, and intuitive. Successful applications of hand gesture recognition include computer games control, human-robot interaction and sign language recognition, to name a few. Vision-based recognition systems can give computers the capability of understanding and responding to hand gestures.

[42] presented analysis of human behaviour in front of a target scene. An application of computer vision techniques is obtain specific information about the behaviour of the people passing in front of a target scene. This is done by analyzing videos captured by cameras monitoring an area under surveillance. The target scene can be a large plasma screen, a projected image an advertising poster or a shop window.

[64] proposed audiovisual information fusion in human–computer interfaces and intelligent environments: A Survey. Microphones and cameras have been extensively used to observe and detect human activity and to facilitate natural modes of interaction between humans and intelligent systems. Human brain processes the audio and video modalities, extracting complementary and robust information from them. Intelligent systems with audiovisual sensors should be capable of achieving similar goals.

[28] presented visual tracking and recognition using appearance-adaptive models in particle filters. The constraints are implemented by including both positive and negative trivial templates in the template set. It dynamically updates the target template set to keep the representative templates throughout the tracking procedure. This is done by adjusting template weights by using the coefficients in the sparse representation. It approaches on five video sequences involving heavy occlusion, large illumination and pose changes.

[68] presented human motion tracking by temporal-spatial local gaussian process experts. Human pose estimation via motion tracking systems can be considered as a regression problem within a discriminative framework. It challenges to mapping from observation space to state space because of the high-dimensional characteristic in the multimodal conditional distribution. In order to build the mapping, existing techniques usually involve a large set of training samples in the learning process which are limited in their capability to deal with multimodality.

[27] presented tracking multiple humans in complex situations. The system parameters are tuned to the characteristics of the scene under analysis. The generative approaches have gained increased attention by the community because they allow overcoming some of the limitations approaches by explicitly modeling the observation process implemented in a camera. The basic idea is to generate a set of likely hypotheses and to compare the actual observation with a

synthetic view of them rendered using explicit models of the different components of the scene such as the targets as well as the background.

[46] presented training hierarchical feed-forward visual recognition models using transfer learning from pseudo-tasks. The convolutional neural networks (CNNs) are a type of deep models in which trainable filters and local neighborhood pooling operations are applied alternatingly on the raw input images, resulting in a hierarchy of increasingly complex features. In addition, CNNs have been shown to be relatively insensitive to certain variations on the inputs.

2.3 REPORTED WORKS ON NEURAL NETWORKS

[2] presented handwritten digit recognition with a back-propagation neural network. Convolutional networks (CN) incorporate constraints and achieve some degree of shift and deformation invariance using three ideas: local receptive fields, shared weights, and spatial sub sampling. The use of shared weights also reduces the number of parameters in the system aiding generalization. Convolutional networks have been successfully applied to character recognition.

[3] proposed convolutional networks for images, speech, and time series. The problem of face recognition from 2D images is typically very ill-posed, i.e. there are many models which fit the training points well, but do not generalize well to unseen images. There are not enough training points in the space created by the input images in order to allow accurate estimation of class probabilities throughout the input space. Additionally, for MLP networks with the 2D images as input, there is no invariance to translation or local deformation of the images.

[20] presented a neural network approach to complete coverage path planning. The objects of interest are extracted from the images by using the dynamic properties (optical flow techniques) and their optical features (color,

texture, shape and so on). The consecutive dynamic behavior of the objects of interest is then estimated based on the current dynamics. By using this predicted dynamics, the amount of data for the region of interest identification can be reduced.

[56] presented classification of fused images using radial basis function neural network for human Face recognition. Fusion of visual and thermal images has been done to take the advantages of thermal images as well as visual images. By employing fusion a new image can be obtained which provides the most detailed, reliable and discriminating information. This method fused images are generated using visual and thermal face images.

[58] proposed neural network based biometric personal identification with fast iris segmentation. The iris recognition system for biometric personal identification is using neural network. It identification consists of localization of the iris region and generation of a data set of iris images followed by iris pattern recognition. In this work, a fast algorithm is proposed for the localization of the inner and outer boundaries of the iris region. Located iris is extracted from an eye image and after normalization and enhancement, represented by a data set.

[76] presented neural networks for face recognition using SOM. It needs to maintain the security of information or physical property is becoming both increasingly important and increasingly difficult. Face recognition is one of the few biometric methods, which is very complicated system since, the human faces change depending on their age, expressions etc. It is not possible to learn all types of expressions into the network.

2.4 REPORTED WORKS ON VIDEO SURVEILLANCE

[4] proposed underwater pipe inspection task using visual serving techniques. It takes a visual serving approach and devises a controller that takes the inputs as the lines extracted from the image of a pope and uses this information to generate steering commands for the ROV Vortex vehicle. A sensor fusion scheme uses the dead reckoning position uncertainty with a 2D position model of the cable to predict the region of interest in the image.

[6] proposed the ability to recognize humans and their activities by vision is key for a machine to interact intelligently and effortlessly with a human-inhabited environment. Because of many potentially important applications, "looking at people" is currently one of the most active application domains in computer vision. This survey identifies a number of promising applications and provides an overview of recent developments in this domain.

[8] proposed W4: real-time surveillance of people and their activities. The visual intelligence of animals can handle by far more complicated tasks, designing an algorithm that works in general circumstances still remains surprisingly difficult. Traditional approaches can work under certain assumptions or when system parameters are tuned to the characteristics of the scene under analysis.

[10] proposed a survey of computer vision-based human motion capture. The focus is based on taxonomy of system functionalities, broken down into four processes: initialization, tracking, poses estimation and recognition. A number of general assumptions used in this research field are identified of these assumptions indicates that the research field is still in an early stage of development.

[11] proposed vision-based control using different cameras for learning the reference image and for servoing. The key idea of the unified approach is to

build a reference in a projective space invariant to camera intrinsic parameters which can be computed if the model is known or if an image of the object is available. Thus, only one low level visual servoing technique must be implemented. The unified visual servoing scheme will be useful especially when a zooming camera is mounted on the end-effector of the robot.

[13] presented the camera mouse: visual tracking of body features to provide computer access for people with severe disabilities. The "Camera Mouse" system has been developed to provide computer access for people with severe disabilities. The system tracks the computer user's movements with a video camera and translates them into the movements of the mouse pointer on the screen. Body features such as the tip of the user's nose or finger can be tracked. The visual tracking algorithm is based on cropping an online template of the tracked feature from the current image frame and testing where this template correlates in the subsequent frame.

[32] presented performance evaluation for video surveillance systems. The tracking systems are evaluated using the ETISEO evaluation tool that defined several metrics for gauging the performance of tracking systems which are split into five groups, Detection, Localization, Tracking, Classification and Event Recognition. Each group of metrics contains several sub-metrics to evaluate specific criteria and a global metric, which is the average of all metrics within the group.

[51] presented recognizing visual focus of attention from head pose in natural meetings. The model is using a Gaussian mixture model (GMM) or a hidden Markov model (HMM) whose hidden states correspond to the VFOA. The potential VFOA of a person is not restricted to other participants only. It includes environmental targets as well (a table and a projection screen), which increases the complexity of the task; with more VFOA targets spread in the pan as well as tilt gaze space.

[53] presented a survey on behavior analysis in video surveillance for homeland security applications. Surveillance cameras are inexpensive and everywhere, but the manpower required to monitor and analyze them is expensive. Consequently the videos from these cameras are usually monitored sparingly or not at all, often used only as archive, to refer back to once an incident is known to have taken place. Surveillance cameras can be a far more useful tool if instead of passively recording footage can be used to detect events requiring attention as they happen, and take action in real time.

[73] presented visual tracking is a fundamental task in computer vision. There has been no systematic way of analyzing visual trackers so far. In this work, methods that can help researchers determine strengths and weaknesses of any visual tracker. To this end, consider visual tracking as an isolated problem and decompose it into fundamental and independent sub problems. Each sub problem is designed to associate with a different tracking circumstance.

[54] proposed hierarchical audio-visual cue integration framework for activity analysis in intelligent meeting rooms. Scene understanding in the context of a smart meeting room involves the extraction of various kinds of cues at different levels of semantic abstraction. Specifically, human activity in a scene is usually monitored using arrays of audio and visual sensors. Tasks such as person localization and tracking, speaker ID, focus of attention detection, speech recognition and affective state recognition are among them.

[59] presented visual servoing set free from image processing. Instead of using geometric features (points, straight lines, pose, homography, etc.) as it is usually done, it uses directly the luminance of all pixels in the image. Since, most of the classical controls laws fail in this case turn the visual servoing problem into an optimization problem leading to a new control law.

[72] presented high-speed vision systems and projectors for real-time perception of the world. The overview of high-speed vision systems that enable

real-time image acquisition and visual processing at frame rates of several hundreds to thousands of frames per second, which are substantially higher than the standard video rates. High-speed vision systems enable fast measurement and control of dynamic systems and have been successfully applied in the fields such as robotics, in which real-time perception of dynamic environment is critically important.

[74] presented visual tracking with singular value particle filter. A robust tracking is an important and challenging problem in computer vision. The existing algorithms do not work well if there are confusing objects in the surrounding environment or the target appearance has a significant change. It describes a novel particle filter for object tracking. It treats the blob image of the object as a matrix and adopt singular values to construct the feature model. In the second stage, the particle filter scheme is applied for tracking.

[69] proposed multi-modal object tracking using dynamic performance metrics. Intelligent surveillance systems typically use a single visual spectrum modality for their input. These systems work well in controlled conditions, but often fail when lighting is poor or environmental effects such as shadows, dust or smoke are present. Thermal spectrum imagery is not as susceptible to environmental effects, however thermal imaging sensors are more sensitive to noise and they are only gray scale, making distinguishing between objects difficult.

[75] proposed multi-target visual tracking based effective surveillance with cooperation of multiple active cameras. This paper presents a tracking-based surveillance system that is capable of tracking multiple moving objects, with almost real-time response, through the effective cooperation of multiple pan-tilt cameras. To construct this surveillance system, the distributed camera agent, this tracks multiple moving objects independently.

2.5 REPORTED WORKS ON HUMAN MOTION DETECTION

[16] presented edges characterize boundaries and are therefore a problem of fundamental importance in image processing. Image Edge detection significantly reduces the amount of data and filters out useless information, while preserving the important structural properties in an image. Since, edge detection is in the forefront of image processing for object detection, it is crucial to have a good understanding of edge detection algorithms.

[43] presented lips detection for audio-visual speech recognition system. Audio-visual speech recognition (AVSR) becomes a research trend due to the stimulation of the restrictions rise from the automatic speech recognition (ASR). With the aid of visual signal, AVSR outperforms ASR under certain undesired circumstances such as noisy environments. The key element for a good performed AVSR is the capability of front end lips detection.

[50] presented edge detection techniques for image segmentation – a survey of soft computing approaches. Soft Computing is an emerging field that consists of complementary elements of fuzzy logic, neural computing and evolutionary computation. Soft computing techniques have found wide applications. An important application is edge detection for image segmentation. The process of partitioning a digital image into multiple regions or sets of pixels is called image segmentation.

[77] presented an approach of face detection using geometrical definition of human face. Human face can be identified by different biometric features, which are genetic properties of a human being. If the biometric features can be extracted from a facial image then they can be used for face recognition. A new method of face detection by using human face geometry. It conducted 25 face images which are captured from CCD camera under the varying condition of illumination, pose and expression.

[31] radar-vision fusion with an application to car-following using an improved AdaBoost detection algorithm. In field of car detection, combination of

haar-like feature and HOG (Histogram of Oriented Gradient) is a way to encode an input image to obtain a vector of visual descriptors. Motion features are used to given the candidate regions and Haar-Like features are used to detect the vehicles. It can significantly increase detection rates for car detection and Haar-Like feature used for zebra detection, real-time road sign detection, in vehicle safety systems.

[34] proposed improving the selection and detection of visual landmarks through object tracking. The unsupervised selection and posterior recognition of visual landmarks is a highly valuable perceptual capability for a mobile robot. It's aims to achieve the capability by combining a bottom-up data driven approach with top-down feedback provided by high level semantic representations. The bottom-up approach is based on three main mechanisms: visual attention, area segmentation, and landmark characterization.

[52] proposed a real-time tracker that simultaneously tracks the 3-D head pose and facial actions in monocular video sequences that can be provided by low quality cameras. It has two main contributions, (i) an automatic 3-D face pose initialization scheme for the real-time tracker by adopting a 2-D face detector and an eigen face system. (ii) it uses the proposed methods—the initialization and tracking—for enhancing the human–machine interaction functionality of an robot.

[65] presented lip detection and tracking using variance based haar-like features and Kalman filter. Speaker lip motion stands out as the most linguistically relevant visual feature for speech recognition. Lip reading is an active field that receives much attention from computer scientists that take part not only in science, such as a speech recognition system, but also in social activities, such as teaching pronunciation for deaf children in order to recover their speaking ability and combining visual features and audio features to increase the accuracy in noisy environments.

2.6 REPORTED WORKS ON HUMAN MOTION ROBUSTNESS IN LEARNING

[7] proposed mobile robot navigation using self similar landmarks. It uses the self-similar gray pattern landmarks for navigation and localization aids. A simple artificial landmark model can be used for the self localization of indoor mobile robots. A pair of color points neighboring each other has been used as a sample to represent the probability density in the Condensation algorithm. The information of a single landmark in a single image presented a localization algorithm to estimate the absolute position accurately.

[14] presented robust and efficient tracking in image sequences using a kalman filter and an affine motion mode. The system is to study the benefits of combining image processing with navigation data that should be available from the control system of any AGV system. It is using a Kalman Filter the tracking algorithm can handle objects large movements between images and it becomes more resistant to the occlusions. However, their tracking system is designed mainly based on the image brightness, which gives much less robust performance under environmental lighting changes.

[24] proposed support vector tracking (SVT) integrates the support vector machine (SVM) classifier into an optic-flow-based tracker. Instead of minimizing an intensity difference function between successive frames, SVT maximizes the SVM classification score. For large motions between successive frames, build pyramids from the support vectors and use a coarse-to-fine approach in the classification stage. It uses SVT for vehicle tracking in image sequences.

[57] proposed iterative soft color-shrinkage for color-image denoising. To remove signal-dependent noise of a digital color camera, present a new soft color-shrinkage scheme for color-image denoising in a wavelet transform domain. The classic soft shrinkage scheme works well for monochrome-image

denoising, to utilize inter-channel color cross-correlations, a noisy image undergoes the color-transformation from the RGB to the luminance-and-chrominance color space and the luminance and the chrominance components are separately denoised.

[66] presented research on preprocessing of palmprint image based on adaptive threshold and euclidian distance. This approach combines the characteristics of palmprint image database with an improved adaptive binary segmentation which is based on gray-scale distribution and structure of the palmprint image. It uses median filter to remove the cross and obscure parts of the breakpoint, take advantage of an improved global threshold algorithm to get the binarization image and carry out corrosion, swelling, edge extraction and boundary tracking.

[29] presented an identification system combined with fingerprint and cryptography. A design schema of a security authentication system combined with fingerprint identification and public key cryptography is explored. In this schema, fingerprint is added into user's private key and served as a security parameter, such that user's secret key is separated into secret key parameters and fingerprint, by secret splitting mechanism, which makes the secret key to be bounded with user's information that increase the security of secret key ultimately.

[67] presented active boosting for interactive object retrieval. In this work, a new algorithm based on boosting for interactive object retrieval in images is propose “online boosting” algorithms where weak classifier sets are iteratively trained from data. These algorithms are for visual tracking in videos and are not well adapted to “online boosting” for interactive retrieval. It used to enhance the strong classifier resulting from “boosting” process, but also to build new weak classifiers.

2.7 REPORTED WORKS ON HUMAN MOTION TRACKING

[21] proposed a multiple-sensor multiple-target tracking approach for the auto-taxi system. The system consists of two basic components: the sensor-level tracking and multiple-sensor track fusion center. Each sensor in the sensor level is considered as an intelligent one which generates its own track files. Thus, the task of the fusion center is to fuse the local track files to produce a more accurate and reliable single system track file.

[33] proposed an adaptive algorithm for eye-gaze-tracking-device calibration. Eye-gaze tracking constitutes a supportive tool to probe the perceptual or the cognitive processes of the subjects. In day-to-day applications, eye-gaze tracking can be used as a computer interface for both industrial and nonindustrial applications, which require hands-free installations. It can also be used to help disabled people to use computers for communication and for environmental control.

[35] presented accurate eye center location and tracking using isophote curvature. The ubiquitous application of eye tracking is precluded by the requirement of dedicated and expensive hardware, such as infrared high definition cameras. Therefore, systems based solely on appearance (i.e. not involving active infrared illumination) are being proposed. However, although these systems are able to successfully locate eyes, their accuracy is significantly lower than commercial eye tracking devices.

[36] presented face tracking and recognition with visual constraints in real-world videos. It track faces using a tracker that adaptively builds a target model reflecting changes in appearance, typical of a video setting. However, adaptive appearance trackers often suffer from drift, a gradual adaptation of the tracker to non-targets. The tracker introduces visual constraints using a combination of generative and discriminative models in a particle filtering framework.

[44] presented tracking facial features under occlusions and recognizing facial expressions in sign language. This work is recognizing facial expressions that are used in sign language recognition. Facial features are tracked to effectively capture temporal visual cues on the signer's face during signing. A Bayesian framework is proposed as a feedback mechanism to the Kanade-Lucas-Tomasi (KLT) tracker for reliably tracking facial features in the presence of head motions and temporary occlusions.

[55] presented the implementation of shadow suppression in intelligent tracking system. It describes the implementation of shadow suppression algorithm in an intelligent tracking system based on multimedia Digital Signal Processor (DSP). In the application, the fast filling and noise elimination algorithm is added to improve the accuracy of shadow suppression. The algorithm level optimization simplifies the complexity of the algorithm.

[63] presented increasing motion information by using universal tracking of 2D-Features. The visual features for recognition are known in advance and remain static due to a controlled environment, object tracking is state-of-the-art. Tracking objects in dynamically changing environments is still a challenge. Even harder is the tracking of moving objects with a moving camera. The algorithm realizes a deterministic approach to track any 2D-features representable in a general way.

[17] presented robust online appearance models for visual tracking. It identified four possible factors causing appearance change, fitting them with a mixture model, which was used to estimate image motion. A more elaborate mixture model fit via an online EM algorithm, in which three components were used to model the responses of wavelet filters and thereby account for appearance variation during tracking. Their method is able to handle variations in pose, illumination and expression.

[70] proposed tracking objects through occlusions using improved Kalman Filter. In a visual surveillance system, robust tracking of moving objects which are partially or even fully occluded is very difficult. In this a paper, a method of tracking objects through occlusions using a combination of Kalman filter and color histogram. By changing covariance of process noise and measurement noise in Kalman filter, this method can maintain the tracking of moving objects before, during and after occlusion.

[71] presented a component-based visual object tracker for mobile platforms. The core of the technique is a component-based descriptor that captures the structure and appearance of a target in a flexible way. It can be learned quickly from a single training image and is easily adaptable to different objects. The descriptor is integrated into the observation model of a visual tracker based on the well known Condensation algorithm.

[25] presented visual tracking and recognition using appearance-adaptive models in particle filters. An approach that incorporates appearance-based models in a particle filter is to realize robust visual tracking and recognition algorithms. In conventional tracking algorithms, the appearance model is either fixed or rapidly changing and the motion model is simply a random walk with fixed noise variance and also, the number of particles is typically fixed.

[26] proposed tracking multiple humans in complex situations. Person detection is performed by splitting the image into sub-regions which contain concentrated areas of motion and then locating heads and fitting ellipses within each region. Working within sub-regions overcomes problems caused by people occupying a common column of the image causing inaccurate vertical projections. This detection process is used for the detection of new tracks and to support the condensation filter tracking.

[40] presented optical flow-based facial feature tracking using prior measurement. The visual tracking is an important research topic in computer

vision and face expression recognition, in which domain-oriented facial features tracking is a very hot spot. In this work, a robust facial feature tracking method takes Lucas-Kanade-Tomasi (KLT) optical flow as basis and corrects the predictions by prior measurement which consists of pupils detecting, feature restricting and errors are accumulating.

2.8 INFERENCE FROM LITERATURE SURVEY

From the literature survey, it is observed that the reported works on dimensionality reduction (feature extraction) for human motion do not perform effectively against the background causes, localization and is less efficient. Also, the previous reported works on human motion estimation does not integrate phase compensators. In this research work, the background subtraction technique is proposed to enhance the human motion tracking and detection. Also, the hardware implementation on STM32 ARM processor is done. The correlation extractor integrates the human motion (video) as input and the correlated output samples are extracted using unsupervised trained weights. The use of Karhumen – Loeve Transform (KLT) has been reported in previous works, as KLT is optimal in terms of compactness of representation. However, KLT requires more pre-processing stages and hence, as an alternative DCT is preferred in this research. Also, DCT closely approximates KLT in the context of information packing.