

ABSTRACT

Driven by key security, safety and commerce related applications; significant increase in demand for embedded vision systems, witnessed in recent years. However, significant research challenges remain as most of these applications do not allow for controlled settings. The common challenge for high level vision algorithms like detection, tracking and recognition is designing object representation techniques which should be robust against low resolution, presence of noise, occlusion, different scale, varying illumination, background clutter and pose change. Another important challenge is to develop implementation strategies for these computational expensive algorithms to make them run in resource constraint embedded processors, which enable low cost vision systems ubiquitous. It is also important to evaluate vision algorithms indented for standalone system, with not only its accuracy but also the computational requirement. Therefore, the two main objectives of this thesis are to investigate the subspace learning based object representation techniques with focus on illumination variation and occlusion and evaluate its real-time performance on embedded systems. In most of the vision based pattern recognition applications, the target appearance model is learned from training images and feature vector of a gray scale image with size $a \times b$ is described as an ab -dimensional vector in an original image space. Moreover, to handle the dynamic variations in uncontrolled settings, robust low-level feature descriptors are applied on training images to obtain the feature vectors. In order to handle the high-dimensional feature space, it is important to learn subspace which can compactly capture the rich image properties. The learned subspace can be used as dimensionality reduction tool and within that any simple classifier can

be applied for classification. First, to evaluate this strategy, we have formulated an object tracking problem within the particle filter framework using partial least squares (PLS) based subspace learning technique. For this binary classification problem, to learn the target, illumination robust high-dimensional feature space is constructed by using multi-scale multi-orientation Gabor wavelet on target and background training samples. The low dimensional discriminative feature subspace is learned by using PLS analysis. Compare to the unsupervised principal component analysis (PCA), PLS can incorporate the target and background class labels in its learning framework which gives more discrimination capability to the model. To validate this we have formulated another binary classification problem to classify the eye state for driver drowsiness detection system. Based on the eye state (open or close), the drowsiness of the driver can be detected by estimating the percentage of eye closure (PERCLOS) metric. For this problem, the eye state classification is done using PLS subspace learning technique. The advantage of PLS over PCA in terms of discrimination strength is shown by plotting the first two factors of the dimensionality reduced two class open and closed eye dataset with the help of support vector machine (SVM) classifier. As better discrimination is achieved using the low dimensional PLS subspace, only 18 support vectors are required to get an optimal hyperplane, whereas using the PCA subspace requires 173 support vectors. Secondly, within the sparse representation framework, PCA based subspace learning technique is investigated for face identification applications. The recent findings in sparse representation based tracking and recognition algorithms strongly prove their robustness against partial occlusion and noise. But, the major problem with conventional l_1 minimization based sparse representation is huge computation time for dictionaries with large feature dimension. By using the subspace learning methods, the training data set can be compactly described and the

orthogonality of the dictionary prototypes (or basis vectors) can be exploited for speedy sparse representation. However, in surveillance kind of applications, new subjects are added incrementally, which necessitates rebuilding the gallery models every time a new subject is added. To address these issues, we have proposed a face identification algorithm using class specific linear subspaces for each subject. The given query image is represented using the fast sparse representation technique against each dictionary. After that, using the representation coefficients, reconstruction error is approximated for each subject to find the best match. In face identification applications, it is essential to reject the invalid test faces often detected in surveillance videos. Therefore, by using the sparse error coefficients, we have defined a metric called error ratio to reject the detector false positives. Finally, to deal with occlusion and illumination variation simultaneously, we have added illumination normalization step in our previously proposed identification algorithm. The lighting variations are minimized by eliminating the low frequency subbands obtained using dual-tree complex wavelet transform (DTCWT). The reason for selection of DTCWT based decomposition is its better directional selectivity and fast computation. It is observed that with this technique, the recognition performance for the challenging test faces affected by both occlusion and illumination variation has been significantly improved. All the proposed algorithms are evaluated using various publically available datasets and their performances are compared with that of the efficient representative methods. Second part of this research is focused on system level implementation of proposed algorithms on embedded video processor (TMS320DM6437). For driver drowsiness detection as well as face identification applications, face detection module plays major role in its performance. For this part, we have chosen Viola-Jones (VJ) detector, due to its robustness and less computation time. Motivated by the observation that non-face search windows are

eliminated in the early stages of cascaded classification, we have optimized VJ detector by using skin color based search window reduction technique. Along with this module, the face identification algorithm is also integrated to evaluate it in real-time surveillance videos for 20 subjects. Moreover, the driver drowsiness detection system is implemented by using the right eye and its performance is validated for normal and drowsy drivers in onboard during day and night using infrared (IR) camera. Finally, to address the alignment issues in video based face identification applications, extended active shape model (ASM) based facial landmarks detection algorithm is investigated elaborately in real-time. The computation time required for key components in each applications are reported in detail and this would suggest the vision system engineers to select the right processing platform for diverse applications.