Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 64 (2013) 159 – 168

Procedia Engineering

www.elsevier.com/locate/procedia

International Conference On DESIGN AND MANUFACTURING, IConDM 2013

Robust object tracking via class aware Partial Least Squares-Gabor Wavelet Subspace

Selvakumar K[a,*] , Jovitha Jerome[b]

[a]*Research scholar, Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore- 641004.*
[b]*Professor, Department of Instrumentation and Control Systems Engineering, PSG College of Technology, Coimbatore- 641004.*

Abstract

Effective tracking is still a big challenge due to lack of robust descriptors which captures discriminative features in non-controlled environment. We propose a novel descriptor based on Gabor wavelet and Partial Least Squares (PLS) discriminant analysis. Multi scale and multi orientation Gabor wavelets can extract selective local frequencies effectively in spatial and frequency domain. Due to the large dimension of feature vectors, dimensionality reduction is done using class aware PLS analysis. Unlike unsupervised Principal Component Analysis (PCA), PLS based subspace model learns target effectively by explicitly knowing the class labels of target and background region feature vectors. Tracking is done using particle filter and similarity between target and candidates is measured using low dimensional subspace model. To combat the target changes during tracking, novel static and dynamic target as well as background update strategy is used. Experimental results of various dataset demonstrate that the proposed tracker improves robustness and accuracy against representative trackers.

Keywords: Object tracking; Gabor wavelet; Partial Least Squares Discriminant Analysis;Particle filter.

1. Introduction

Even though automated video surveillance is being researched for last one decade, their success in laboratory still unsuitable for use in real-world scenarios [1]. On other hand, it is impossible to deploy human operator for all cameras and even in case of deployment, after continuous monitoring for 12 minutes and 22 minutes an operator likely to miss 45% and 95% of screen activity respectively [2]. However, the usable real-time intelligence gathered automatically from huge video data can help law enforcement and military to play proactive role. In any video surveillance system, detection and tracking modules play a crucial role and which should be robust against illumination change, distortion, pose change, cluttered background and occlusion. By initializing the target object in first frame manually, this paper focuses only on visual object tracking. Entire visual object tracking history can be found in [3] [4]. Generally success of any tracker lies in extracting promising discriminate features of the target. Wavelet filter bank analysis provides selective frequency components by omitting other noise components for different orientations [5].In general, Gabor wavelet features are proven to be robust in various computer vision tasks like texture analysis, detection, recognition, image retrieval and classification [6][7]. But

* Corresponding author. Tel.:+91-9095252193; fax: +0-000-000-0000 .
E-mail address: kskumareee@gmail.com

doi:10.1016/j.proeng.2013.09.087

Gabor feature space is very large dimension and it cannot be directly used for any application which demands real-time performance. For example, sparse representation based trackers [8][9] take huge computation time for large dimension underdetermined bases matrix. In many applications, PCA based dimensionality reduction technique is being used successfully to learn target from large dimension feature vectors [10][11]. But main problem with PCA is that, the chosen principal components are having no relevance with response variables. Unlike detection and recognition, in tracking, prior training samples are not available. David Ross et al. [12] proposed PCA based incremental learning for visual tracking, which uses Eigen subspace to model the target. Wenli et al. [13] proposed multiple cue subspace learning method. Ming Che et al. [14] also proposed PCA based region-wise linear subspace method, which partition the target appearance into several sub images for linear subspace representation. However, in specific to tracking, due to its unsupervised nature of learning, negative background samples cannot be used to learn target, which generally improves discrimination between target and background. Clearly for discrimination as well as dimension reduction PLS is preferred over PCA [15]. Larry Davis et al. [16][17] proposed PLS based dimensionality reduction for human and vehicle detection applications and this method demonstrates promising results . The supervised nature of learning target by knowing the target and background class labels shows its suitability for formulation of tracking framework. Ordinary least squares regression cannot be used due to multicollinearity problem. Recently, Ming-Hsuan et al. [18] proposed PLS based multiple appearance model tracker with direct pixel intensity as feature vector. But this method suffers from drift in many test cases with drastic illumination variation and heavy occlusion. Even for gradual illumination variation, this tracker demands large number of particles which leads to heavy computation cost. Based on above analysis, we propose PLS- Gabor Wavelet subspace based tracking framework, which improves tracker performance in following ways; first multiscale multi orientation Gabor wavelet features are extracted which are robust against illumination variation. Then better discrimination ability of the tracker between target and background is achieved using PLS method by knowing target and background class labels. Finally to combat gradual pose change of the target during tracking, novel target and background update strategy is used. Bayesian inference based particle filter method is used to estimate target in each frame using the similarity measure [19].

2. Particle filter framework

For tracking, the state s_t of the target in current frame is estimated within the Bayesian framework. Particle filters provide the framework to estimate and propagate the non-Gaussian posterior probability density function of state variables. The predicting distribution of s_t for given all available observations z_{t-1} till time $t-1$ is given as;

$$p(s_t \mid \mathrm{z}_{1:t-1}) = \int p(s_t \mid \mathrm{s}_{1:t-1}) p(s_{t-1} \mid \mathrm{z}_{1:t-1}) dx_{t-1} \tag{1}$$

At time t for given all observation $z_{1:t}$, the posterior distribution obtained using Bayes theorem is

$$p(s_t \mid \mathrm{z}_{1:t}) \alpha\, p(z_t \mid \mathrm{s}_t) \int p(s_t \mid \mathrm{s}_{1:t-1}) p(s_{t-1} \mid \mathrm{z}_{1:t-1}) dx_{t-1} \tag{2}$$

where $p(z_t \mid \mathrm{s}_t)$ is observation likelihood model. This distribution can be approximated by using finite number of samples $\{s_t^i \mid i = 1,...,N_s\}$. These samples are generated in consecutive frames using transition distribution $p(s_t \mid \mathrm{s}_{t-1})$. The target state at time t is defined by $s_t = (x_t, \mathrm{y}_t, \eta_t, \mathrm{s}_t, \beta_t, \phi_t)$ where $x_t, \mathrm{y}_t, \eta_t, \mathrm{s}_t, \beta_t, \phi_t$ are x, y translations, rotation angle, scale, aspect ratio and skew direction respectively. The parameters in the state are modeled independently with a Gaussian distribution based on its previous state parameters s_{t-1} . The state transition distribution is

$$p(s_t \mid \mathrm{s}_{1:t-1}) = \mathcal{N}(s_t; s_{t-1,} \Sigma) \tag{3}$$

Where Σ is a diagonal covariance matrix, whose elements are the corresponding variances of the affine transform parameters. In this work $p(z_t \mid s_t)$ is obtained based on residual error (d_i) i.e., similarity between target and candidates models approximated using PLS-Gabor wavelet subspace. For given set of samples in current frame observation likelihood can be calculated by

$$p(z_t \mid \mathrm{s}_t) \alpha \exp(-d_i) \tag{4}$$

The *Maximum A Posteriori* (MAP) estimation in current frame for given N_s samples is estimated by

$$\hat{s}_t = \operatorname{argmax} p(s_t \mid \mathrm{z}_{1:t}) \tag{5}$$

Then estimated target location is further used to draw the samples in next frame and this process continues for all frames.

3. Formulation of PLS-Gabor Wavelet Subspace

Target appearance model is formulated using Gabor wavelet feature space and PLS discriminant analysis. Methodology involved in construction of low dimension subspace is given below.

3.1 Gabor feature extraction

This section describes motivation behind extracting Gabor features of the target and background templates rather than using pixel intensity directly or other descriptors. It is very essential to describe a target robust against major tracking constraints like illumination change, distortion, cluttering environment, small pose change etc. Biological motivation behind the Gabor wavelet is that, it can best model human visual system. Gabor filter can effectively extract selective local frequency components in different orientations when compare to other spatial descriptors. The Gabor filters are defined as in [5]

$$\Psi_k\ l\ = \frac{k^2}{\sigma^2}\exp\left(-\frac{k^2}{2\sigma^2}l^2\right)\left(\exp\ ikl\ -\exp\left(-\frac{\sigma^2}{2}\right)\right) \tag{6}$$

where l is the variable in spatial domain and k is the frequency vector which determines scale and orientation. In this paper, two scale and eight orientations Gabor kernels are used to extract the features as shown in Fig 1. Each positive (target) $a^p \in \Re^{32\times32}$ and negative (background) template $a^n \in \Re^{32\times32}$ is decomposed into sixteen Gabor transformed templates using convolution as given below.

$$G_k(x,y) = a(x,y) * \psi(x,y) \tag{7}$$

Feature vector $x \in \Re^{1\times m}$ is formed by concatenating all sixteen Gabor transformed templates. Since the feature vector dimension m is very large (in this work, 16384), it cannot be directly used to represent the target. To create feature subspace, dimensionality reduction is done using PLS regression analysis.

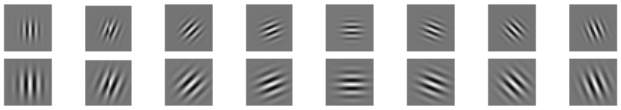


Fig. 1. Real part of Gabor 32×32 kernels for two scale and eight orientations

3.2 Partial Least Squares dimensionality reduction

In PLS analysis, partial least squares discriminant analysis (PLS-DA) is a variant, when response matrix takes only binary values. Detailed information about PLS regression analysis can be found in [20]. A brief overview with respect to proposed tracking framework is given below. Let $X \in \Re^{N\times m}$ be predictor matrix, denotes N feature vectors obtained from positive and negative samples with dimension m and $Y \in \Re^{N\times1}$ be response matrix, and denotes class labels of positive and negative feature vectors. For an input frame, feature vectors represents positive and negative samples are assigned as class label '1' and '0' respectively as shown in Fig 2. Both mean centered X and Y are decomposed using PLS regression analysis as following;

$$\bar{X} = TP^T + E\ ;\ \bar{Y} = UQ^T + F \tag{8}$$

where $E \in \Re^{N\times m}$ and $F \in \Re^{N\times1}$ are predictor and response residuals respectively. $P \in \Re^{m\times p}$ and $Q \in \Re^{1\times p}$ are loading matrices. $T \in \Re^{N\times p}$ and $U \in \Re^{N\times p}$ are latent feature matrices obtained using p PLS weight vectors $w \in \Re^m$, as shown below;

$$T = \bar{X}W\ ; \tag{9}$$

$$W = \left[w_1,...,w_p\right] \in \Re^{m\times p} \tag{10}$$

W is obtained using non-linear iterative partial least squares (NIPALS) algorithm [17], such that residuals are very small for p weight vectors. Amount of cumulative variances of first p weight vectors explained using response variables is shown in Fig 3. In this work p is set to 6 by trial and error method. This constructs new predictor matrix which has large covariance with response variables and it gives better discriminative strength to the target model. Target appearance is modelled by projecting a mean centered positive sample feature vector $\bar{x} \in \Re^{1\times m}$ on W. This gives low dimension subspace model as

$$h = W^T \bar{x}^T \in \Re^p \tag{11}$$

Then within this subspace, similarity measure is constructed to evaluate distance between learned target appearance model and S number of candidate's model. Let $C \in \Re^{S \times m}$ be the feature matrix for S candidates, obtained using particle filter for every frame. This matrix is projected onto W to obtain S candidate models. The similarity between target and candidates are calculated as given below;

$$d_i = \left\| h - W^T \bar{C}_i^T \right\|_2^2 ; i = 1,.....,S \tag{12}$$

$\bar{C}$ is mean centered candidate feature vector matrix. By using Eq. (5), target location of current frame is estimated.

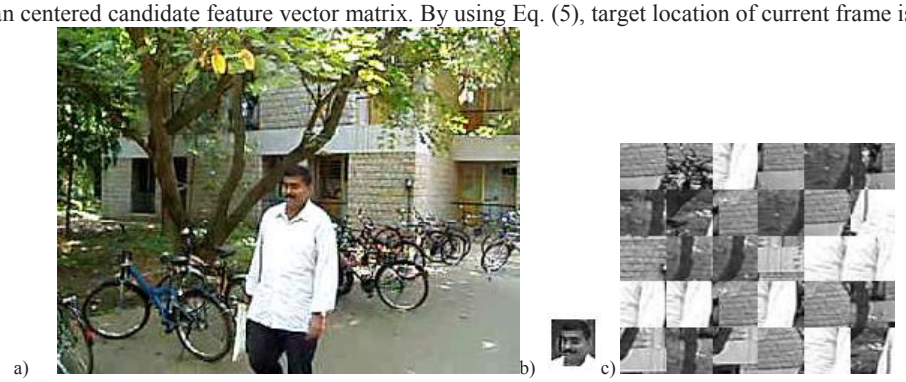


Fig.2. Target learning in initial frame a) Input 240×320; b) One Positive Sample 32×32: Class 1; c) Thirty Negative Samples 32×32×30: Class 0

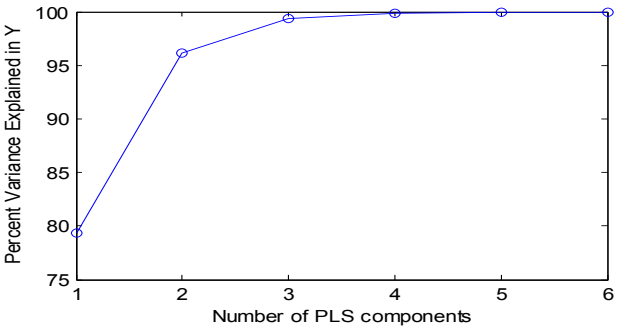


Fig. 3. Cumulative variance of first six PLS components

3.3 Dynamic model update strategy

During the tracking process, target and background appearance undergo various changes. Most of the tracking frameworks use single linear target model and very recently multiple target model is proposed [18]. All these frameworks update their target models during tracking. But, major problem with these update strategies is original target model obtained in initial frame will be lost if target is updated with background noise or during occlusion. In the proposed method, multiple appearance models are used as shown below.

$$H = \{h_j\}; j = 1,...,J \tag{13}$$

where h is target appearance model and J is predefined number of appearance models. But in this method, reliable original model h_1 obtained from initial frame remains unchanged till complete track span. This helps tracker to sustain even for long duration partial occlusion. Remaining models are updated as following; If estimated candidate's similarity is less than predefined threshold and if already more than one model exists $(j > 1)$, gradual changes on the target are updated on model with maximum similarity as given below;

$$a^p = fa^p + (1-f)\mathrm{a}_t^p \tag{14}$$

where f is forgetting factor, a^p is positive template of that appearance model, a_t^p is estimated positive template of that frame. Using updated a^p and current background information, corresponding model h_j is updated using new W_j obtained

by PLS analysis. Otherwise, new model h_{j+1} is added into H using a_t^p and current background information as shown in Fig 4. If number of models exceeds predefined number J $(j > J)$, model with minimum similarity is replaced with new model. All the steps involved in PLS-Gabor wavelet subspace based tracking algorithm are given in algorithm 1.

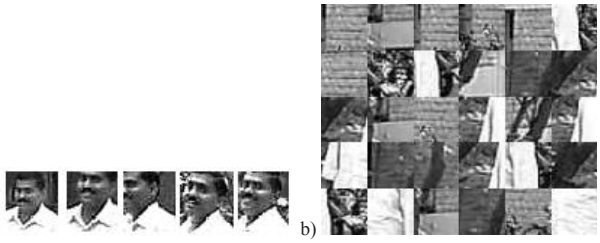


Fig. 4. Online target learning strategy (a) Multiple appearance model positive templates; (b) Current negative templates

Algorithm 1: Proposed PLS-Gabor wavelet subspace based tracking framework

Input: Video frames t, Position of the target (x, y) in first frame, forgetting factor f, Number of target candidates S, Number of positive templates, Number of negative templates, Maximum number of appearance models J.

1: Obtain original target appearance model h_1 using positive and negative templates by PLS regression analysis
2: **for** t=2: last frame
3: Extract S candidate templates
4: Obtain candidates Gabor feature vector matrix C using Eq. (7)
5: Project candidates feature vector on $W_{1:j}$ to get candidate's model
6: Compute similarity measure using Eq. (12) for all target appearance models
7: Estimate target location in current frame using Eq. (5)
8: **if** $d_t^j < Threshold$ (e.g., $\frac{1}{5}\left\|h_j\right\|^2$) and $j > 1$ **then**
9: Update a^p of maximum similarity model using f as given in (14)
10: Update W_j and corresponding model h_j with new a^p and current frame background using step 1
11: **else**
12: **if** $j < J$ **then**
13: Add new model h_{j+1} with a_t^p and current frame background using step 1
14: **else**
15: Learn new model with a_t^p and current frame background using step 1
16: Replace model with maximum error by new learned model
17: **end if**
18: **end if**
19: **end for**

4. Experiment results and Discussion

The proposed tracker is implemented in MATLAB and tested on various public video sequences using 2.53 GHz Intel Core 2 Duo processor. For PLS regression analysis number of weight vectors is set to 6. The Proposed tracker performance is compared with partial least squares tracker (PLS) [18] and incremental subspace tracking (IVT) [12] methods. All these trackers use same dynamic model for particle filtering, but proposed method uses only 30 particles and other two uses 600 particles. For all trackers, wrapped templates are normalized to 32x32 sizes. IVT tracker does not use background information to learn target model. The proposed tracker takes about 0.9 second to process a frame.

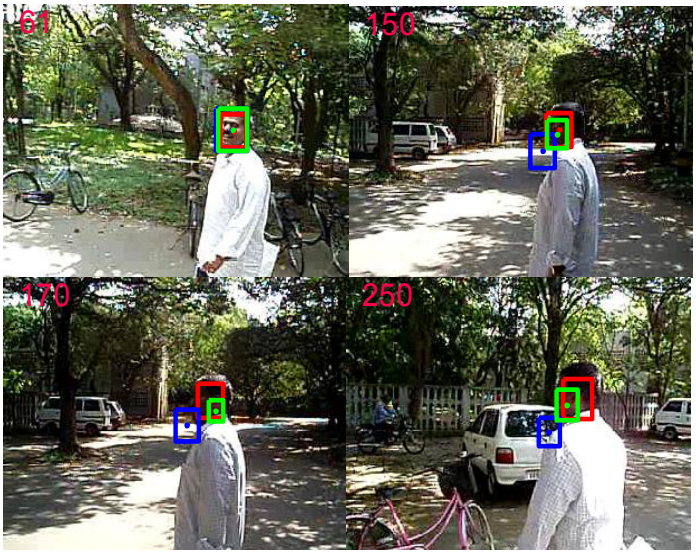
Experiments are carried out to validate the proposed method using videos with different tracking constraints in non-controlled tracking environment. In *face_illumination* sequence, face undergoes drastic illumination variation and gradual pose change with initial cluttering. In *PkTest02* infrared sequence, car is occluded by trees and illumination variation due to shadows. The *faceocc1* and *faceocc2* sequences are particularly used to check robustness against long duration partial occlusions. In *trellis* sequence, face undergoes heavy illumination variation and gradual pose change. In *WomenSequence*,

women undergo long term partial occlusion and pose change. Qualitative results of all video sequences are given in Fig 5. It can be seen that proposed method tracks the target robustly against other two methods for complete track span in all sequences. For quantitative comparisons, the tracking accuracy is measured based on position error. The position error is defined as the distance between estimated location of the tracker and the manually labelled ground truth. Ideal position error should be around zero. Fig. 6 shows position error plots for all test sequences. Using position errors of complete track span, mean error is calculated as

$$ME = \frac{\sqrt{\sum_{t=1}^{T} (GT_t - EL_t)^2}}{T} \tag{15}$$

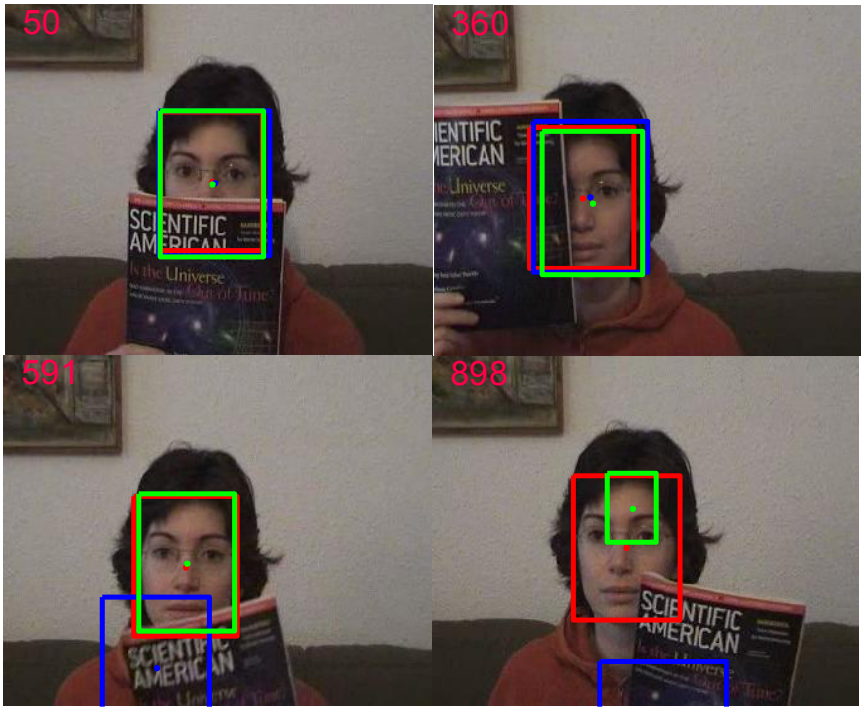
where ME , GT , EL and T are mean error, ground truth, estimated target location and total number of frames respectively. Table 1 summarizes the performance of the representative trakers. It can be seen that proposed method outperforms other two methods in most of the test sequences. Even though IVT done well in face_illumination and faceocc1 sequences, Fig 5.a and 5.c shows that its success rate (percentage of overlap between tracked bounding box and ground truth bounding box) is very low.

a)

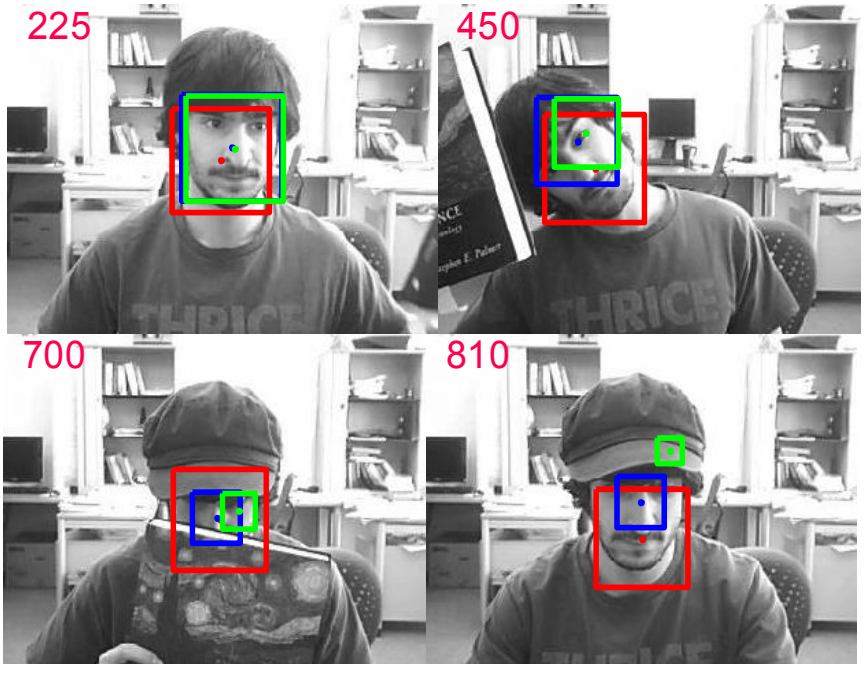


b)

c)



d)



e)

f)

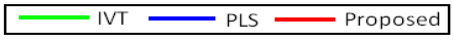


Fig 5. Tracking results of test video sequences a) *face_illumination;* b) *Pk_Test02;* c*) faceocc01;* d) *faceocc02;* e) *WomenSequence;* f) *trellis*

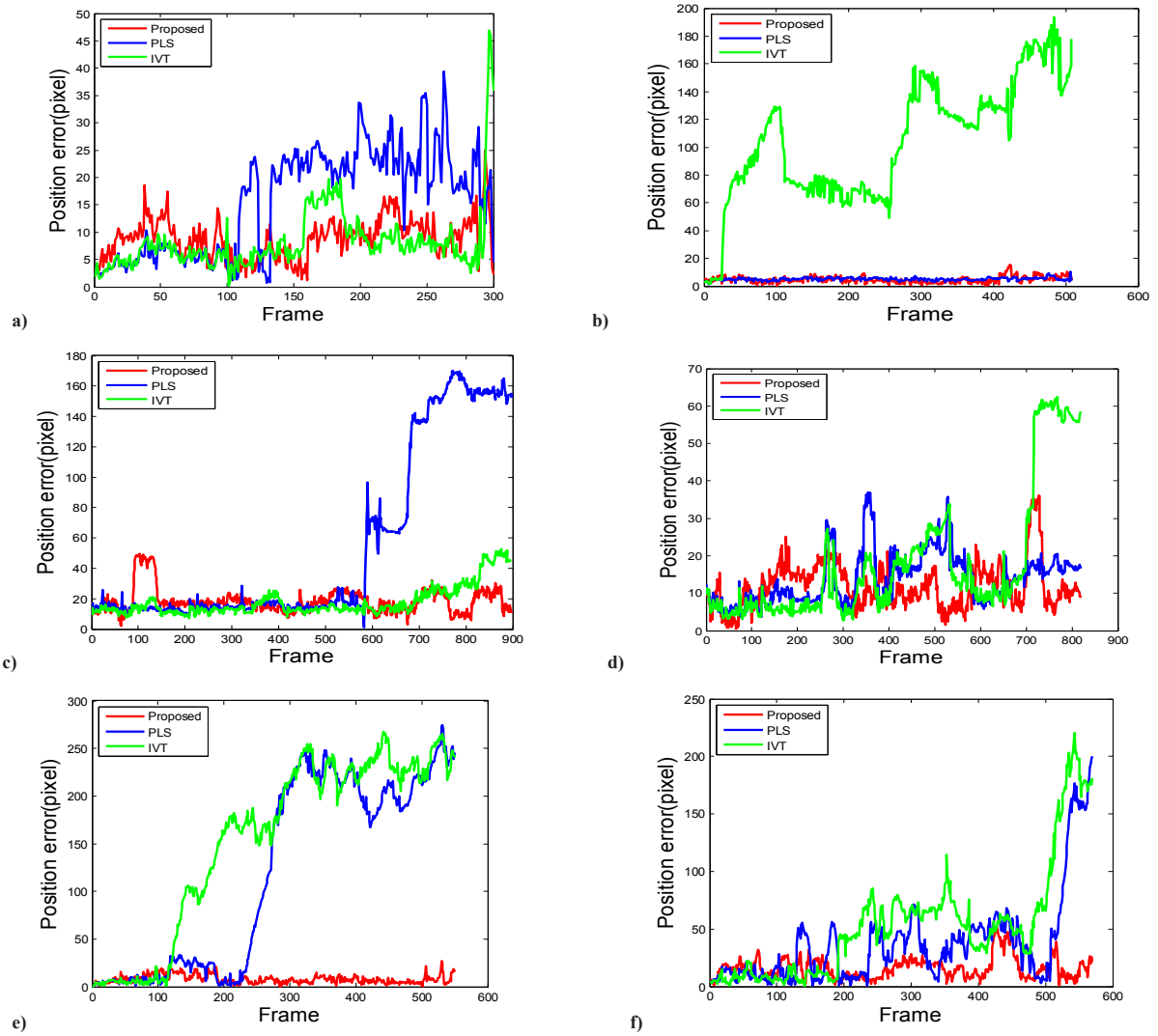


Fig. 6. Position error plots of test sequences a) *face_illumination;* b) *Pk_Test02;* c) *faceocc01;* d) *faceocc02;* e) *WomenSequence;* f) *trellis*

Tables 1. Mean errors of test video sequences

Methods	Mean error (in pixels)					
	Face_illumination	*Pk_Test02*	*faceocc1*	*faceocc2*	*WomenSequence*	*trellis*
IVT	8	104	17	18	153	53
PLS	15	5	54	13	119	37
PLS+Gabor	**8**	**4**	18	**11**	**7**	**15**

5. Conclusion

In this paper, we proposed a PLS-Gabor wavelet subspace for robust tracking under non-controlled environments. The proposed tracker achieves better robustness due to multiscale multi orientation Gabor wavelet features, class aware PLS subspace learning method and novel online target and background update strategy. Compare to conventional subspace

learning methods, PLS method knows target and background class labels explicitly, which in turn improves the discrimination power of the target model. Experimental results on various test sequences show that the proposed tracker achieves lower tracking error with reasonable computation time when compare to other representative trackers. In future work, focus will be on constructing application specific tracking framework rather than general one.

References

[1] Dadashi, N., Stedmon, A, W., and Pridmore, T. P., 2012. Semi-automated CCTV surveillance: The effects of system confidence, system accuracy and task complexity on operator vigilance, reliance and workload, Applied Ergonomics, pp. 1-9.

[2] Anisworth, T., 2002. Buyer beware, Security Oz, 19, pp. 18-26.

[3] Yilmaz, A., Javed, O., and Shah, M., 2006. Object tracking: A survey, ACM Computing Survey, 38, (4).

[4] Porikli, F., and Yilmaz, A., 2012. Object Detection and Tracking, Video Analytics for Business Intelligence, Studies in Computational Intelligence, 409, pp. 3-41.

[5] Kyrki, V., Kamarainen, T. K., and Kalviainen, H., 2006. Invariance Properties of Gabor Filter-Based Features- Overview and Applications, IEEE Transactions on Image Processing, 15, (5), pp. 1088-1099.

[6] Ting, R., Zhang, Q., Zhou, Y., and Xing, J., 2013. Object tracking using particle filter in the wavelet subspace, Journal of Nurocomputing.

[7] Cristina, C., Daniela, M., De Diego, I.M., and Cabello, E., 2013. HoGG: Gabor and HOG- based human detection for surveillance in non-controlled environments, Journal of Nurocomputing, 100, pp. 19-30.

[8] Mei, X., and Ling, H., 2009. Robust visual tracking using l1 minimization, Proc. Of IEEE International Conference on Computer Vision, Kyoto, Japan, pp. 1436-1443.

[9] Mei, X., and Ling, H., 2011. Robust visual tracking and vehicle classification via spare representation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33, (11), pp. 2259-2272.

[10] Yuela P.C., Dao, Q., and Guo-can, C., 1998. Wavelet based PCA for human face recognition, IEEE Southwest Symposium on Image Analysis and Interpretation, Arizona, USA, pp. 223-228.

[11] Fernando, D.L.T., and Michael, J.B., 2003. A framework for robust subspace learning, Journal of Computer Vision, 54, (1-3), pp. 117-142.

[12] David A. R., Jongwoo, L., Ruei-Sung, L., and Ming Hsuan, Y., 2008. Incremental Learning for Robust Visual Tracking, International Journal of Computer Vision, 77, (1-3), pp. 125-141.

[13] Wang, Q., Feng, C., and Wenli, X., 2011. Adaptive multi-cue tracking by online appearance learning, Journal of Nurocomputing, 74, pp. 1035-1045.

[14] Ming-Che, H., Cheng-Chin, C., and Yang, Y.S., 2012. Object tracking by exploiting adaptive region-wise linear subspace representations and adaptive templates in an iterative particle filter, Pattern Recognition Letters, 33, pp. 500-512.

[15] Barker, M., William, R., 2003. Partial Least Squares for Discrimination, Journal of Chemometrics, pp. 166-173.

[16] William, R.S., Aniruddha, K., David, H., and Larry, S. D., 2009. Human Detection Using Partial Least Squares, 12th IEEE International Conference on Computer Vision, Kyoto, Japan, pp. 24-31.

[17] Aniruddha, K., David, H., and Larry, S. D., 2011. Vehicle detection using partial least squares, IEEE Transactions on Pattern Analysis and Machine Intelligence, 33,(6), pp. 1250-1265.

[18] Qing, W., Feng, C., Wenli, X., and Ming-Hsuan, Y., 2012. Object tracking via Partial Least Squares Analysis, IEEE Transactions on Image Processing, 21, (10), pp. 4454-4465.

[19] Nummiaro, K., Meier, E., and Gool, L., 2003. An Adaptive Color Based Particle Filter, Image and Vision Computing, 21, (1), pp.99-110.

[20] Geladi, P., and Kowalski, B., 1986. Partial Least Squares Regression: A Tutorial, Analytica Chimica Acta, 185, pp. 1-17.