Incremental Non-negative Matrix Factorization for Dynamic Background Modelling

Serhat S. Bucak[1], Bilge Gunsel[1] and Ozan Gursoy[1]

[1] Multimedia Signal Processing and Pattern Recognition Lab. Dept. of Electronics and Comm. Eng. Istanbul Technical University 34469 Maslak Istanbul, Turkey

Abstract. In this paper, an incremental algorithm which is derived from Non-negative Matrix Factorization (NMF) is proposed for background modeling in surveillance type of video sequences. The adopted algorithm, which is called as Incremental NMF (INMF), is capable of modeling dynamic content of the surveillance video and controlling contribution of the subsequent observations to the existing representation properly. INMF preserves additive, parts-based representation, and dimension reduction capability of NMF without increasing the computational load. Test results are reported to compare background modeling performances of batch-mode and incremental NMF in surveillance type of video. Moreover, test results obtained by the incremental PCA are also given for comparison purposes. It is shown that INMF outperforms the conventional batch-mode NMF in all aspects of dynamic background modeling. Although object tracking performance of INMF and the incremental PCA are comparable, INMF is much more robust to illumination changes.

1 Introduction

Automatic visual tracking in surveillance video sequences has been an important research area. The fundamental step of this problem is modeling the statistical properties of background successfully and adapting the background representation to content changes experienced in the latter stages of the video scene [1].

Background modeling of samples obtained from outdoor surveillance video sequences may be more challenging as the illumination is very likely to change throughout the scene. Moreover, the variety among the semantic features of scene objects, such as their size, relative motion, occlusion, etc. make the problem even harder. Therefore, the background modeling algorithm should be robust against distortions caused by illumination, as well as having the ability of adapting to dynamic background changes and modifying the former background representation according to content changes. What is meant by a dynamic change is entrance/leaving of an object into/from the scene or changes in object's motions. For instance, when a mobile object in the scene stops, the algorithm should integrate that object into the

This work is partially supported by the Scientific and Technological Research Council of Turkey (TUBITAK) BIDEP

Gunsel B., S. Bucak S. and Gursoy O. (2007).
Incremental Non-negative Matrix Factorization for Dynamic Background Modelling.
In *Proceedings of the 7th International Workshop on Pattern Recognition in Information Systems*, pages 107-116
DOI: 10.5220/0002425501070116

background model as soon as possible. In contrast, a formerly stable object should be treated as a foreground object immediately after it moves.

Nonnegative Matrix Factorization (NMF), with its ability to reduce dimension and extract intuitive features in an efficient and simple way, is a powerful decomposition technique. Furthermore, its constraint of non-negativity makes NMF an intuitive, parts-based representation by allowing only additive combinations of the basis vectors [2], [3]. This is why NMF attracted interest of researchers in several applications including face recognition [4], and biomedical applications [3].

NMF's prior success in revealing latent features in data and its dimension reduction capability makes it a hot prospect for video applications. Thus, we propose usage of NMF for the statistical modeling of background in surveillance video. However, the conventional NMF with its batch nature is not suitable for video content representation. Therefore, in [5] an incremental NMF algorithm (INMF) which is suitable to video analysis is introduced. In this paper INMF is adopted to the statistical background modeling problem and an on-line algorithm which allows dynamic updating of the background model in surveillance video is derived.

In the literature, there is a number of work on statistical modeling of the background. In [6], incremental principal component analysis (IPCA) is proposed for dynamic background modeling. In addition, there are also other algorithms which use the batch-mode PCA [7], or robust PCA [6] for the background modeling. As of our knowledge, there is no reported work that uses NMF with the same objective.

The paper is organized as follows: Necessary mathematical definitions and difficulties with the conventional NMF are given in section 2. In Section 3, the incremental NMF is described. After summarizing usage of the incremental PCA algorithm for background modeling in Section 4, test results are reported in Section 5. The final remarks are given in Section 6.

2 The Conventional Non-negative Matrix Factorization

2.1 Mathematical Definitions

The aim of non-negative matrix factorization (NMF), with rank r, is to decompose the data matrix $\mathbf{V} \in \mathrm{R}^{n \times m}$ into two matrices; which are $\mathbf{W} \in \mathrm{R}^{n \times r}$, also called as the mixing matrix, and $\mathbf{H} \in \mathrm{R}^{r \times m}$, named as the encoding matrix [2],[3].

$$\mathbf{V} \approx \mathbf{WH} \tag{1}$$

As it is formulated in Eq. (1), NMF aims to find an approximate factorization that minimizes the reconstruction error. Different cost functions based on the reconstruction error have been defined in the literature, but because of its simplicity and effectiveness, the squared error given in Eq. (2) is used in this work.

$$F = \| \mathbf{V} - \mathbf{WH} \|^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} \left(V_{ij} - (\mathbf{WH})_{ij} \right)^2 , \tag{2}$$

where subscription ij stands for the ij^{th} matrix entity.

In order to minimize the mean squared error F, which is a convex function of $\mathbf{W}$ and $\mathbf{H}$ separately, Lee and Seung offered the multiplicative update rules given in Eq. (3), where t refers to the iteration number, T denotes the transpose, $a = 1,2,\ldots,r$; $i = 1,2,\ldots,n$, and $j = 1,2,\ldots,m$.

$$H_{aj}^{t+1} = H_{aj}^{t} \frac{\left(\mathbf{W}^{t^T}\mathbf{V}\right)_{aj}}{\left(\mathbf{W}^{t^T}\mathbf{W}^{t}\mathbf{H}^{t}\right)_{aj}} \quad , \quad W_{ia}^{t+1} = W_{ia}^{t} \frac{\left(\mathbf{V}\mathbf{H}^{t+1^T}\right)_{ia}}{\left(\mathbf{W}^{t}\mathbf{H}^{t+1}\mathbf{H}^{t+1^T}\right)_{ia}} \quad . \tag{3}$$

2.2 Difficulties with the Conventional NMF

By offering dimension reduction as well as giving intuitive, additive and parts-based representations of the data, NMF can be considered as an efficient method for video processing. However, the conventional NMF requires re-execution of the algorithm repeatedly as each new frame arrives, if the background representation is to be updated. The effect of this on computational complexity has two aspects. Firstly, as new frames are gathered, the rank of the data matrix $\mathbf{V}$ and correspondingly the rank of the encoding matrix $\mathbf{H}$ increase, causing an increase in the number of update operations per iteration. Secondly, bigger ranks for matrices $\mathbf{V}$ and $\mathbf{H}$ will obviously increase the computational load, as there are matrix multiplications in the update formulas. Besides, as it is shown in Eq.(3), storing the matrix $\mathbf{V}$ is a necessity for batch-mode NMF, since $\mathbf{V}$ is used in update operations of both $\mathbf{W}$ and $\mathbf{H}$. This requirement is another reason that makes batch NMF impractical for video processing. Therefore, a proper incremental NMF algorithm which is able to update the previous representations of video according to the last arrived frame without causing a heavy workload is introduced in [5].

Regarding the background modeling problem in surveillance video, in the batch-mode NMF, the effect of each sample (frame) on the representation is the same, which may cause a difficulty in tracking the dynamic content changes throughout the scene. This is because an efficient background modeling scheme should be capable of assigning higher weights to the recent frames while it reduces the effect of old frames in the representation properly. Therefore, in this paper we propose a scheme that adopts INMF algorithm [5] to the dynamic background modeling problem. It is achieved by deriving an exponential weighting scheme which allows timely tracking the dynamic background changes. The proposed algorithm is presented in the next section.

3 Dynamic Background Modeling by Incremental NMF

The background modeling scheme should be able to make the representation adaptive for content changes, without increasing the computation load. Thus, an incremental-mode algorithm that updates the current representation as each new frame is received would answer the requirements. In the following paragraphs, we describe the

proposed algorithm which adopts incremental NMF [5] to the background modeling problem.

Since the data matrix is constructed by cascading the frames, a new frame will add a new column to both the matrices **V** and **H** shown in Eq. (1). Moreover, in each step, the mixing matrix **W** should be updated with the contribution of the new frame. To achieve this, first of all, effect of the new frame (sample) on the cost function should be examined.

Let F defined in Eq.(2) be the cost function of m frames; thus is denoted as F_m. Similarly, the matrices **V**, **W** and **H** shown in Eq.(2) which are calculated for the first m frames are denoted by $\mathbf{V}_m$, $\mathbf{W}_m$ and $\mathbf{H}_m$, respectively. As a new sample ($(m+1)^{th}$ frame) **v** arrives, a new component that formulizes reconstruction error of **v** is added to the cost function as it is shown in Eq.(4). In Eq.(4) v_i refers the i^{th} element of **v** and h_a denotes the a^{th} component of **h**, which is the new column of the encoding matrix. In Eq.(4) we introduce a parameter, α, which is crucial in controlling the algorithm's ability to adapt to dynamic content changes. α can take any value in the interval $(0,1)$.

$$F_{m+1} = (1-\alpha)F_m + \alpha\sum_{i=1}^{n}\left(v_i - \sum_{a=1}^{r} W_{ia}h_a\right)^2 . \tag{4}$$

In order to obtain a NMF representation for the new data matrix $\mathbf{V}_{m+1} \in \mathrm{R}^{n\times(m+1)}$, we need to minimize F_{m+1} with respect to $\mathbf{W}_{m+1}$ and $\mathbf{H}_{m+1}$. Since the cost function F_{m+1} defined in Eq.(4) is a convex function of $\mathbf{W}_{m+1}$ and $\mathbf{H}_{m+1}$ separately, as it is used for the conventional NMF [3], we can use the gradient descent algorithm in the optimization. Note that each frame in $\mathbf{V}_{m+1}$ is reconstructed by the help of the corresponding column of the encoding matrix $\mathbf{H}_{m+1}$, thus we just need to take the derivatives with respect to h_a, and W_{ia} which refers to the ia^{th} entity of the mixing matrix $\mathbf{W}_{m+1}$. Taking the partial derivatives and choosing a proper step size yields the update rules given in Eq.(5) [5].

$$h_a^{t+1} = h_a^t \frac{\left(\mathbf{W}_m^{t\,T}\mathbf{v}\right)_a}{\left(\mathbf{W}_m^{t\,T}\mathbf{W}_m^t\mathbf{h}^t\right)_a}, \quad W_{ia}^{t+1} = W_{ia}^t \frac{\left((1-\alpha)\mathbf{V}_m\mathbf{H}_m^T + \alpha\mathbf{v}\mathbf{h}^{t+1^T}\right)_{ia}}{\left(\mathbf{W}_m^t\left((1-\alpha)\mathbf{H}_m\mathbf{H}_m^T\right) + \alpha\mathbf{h}^{t+1}\mathbf{h}^{t+1^T}\right)_{ia}} \tag{5}$$

Note that, unlike the conventional NMF that requires updating all the elements of $\mathbf{W}_{m+1}$ and $\mathbf{H}_{m+1}$, whenever the $(m+1)^{\text{th}}$ frame arrives, INMF does not need to update all elements of the encoding matrix $\mathbf{H}_{m+1}$ for the previous frames, but only the components corresponding to the new frame are updated. As a result, the number of updating per iteration is fixed, that significantly reduces the computational complexity. Furthermore, since the matrices $\mathbf{V}_m$ and $\mathbf{H}_m$ remain the same throughout the iterations, the algorithm computes the multiplications $\mathbf{V}_m\mathbf{H}_m^T$ and $\mathbf{H}_m\mathbf{H}_m^T$ once, which also reduces the complexity. Update iterations are repeated till convergence

and the basis matrix $\mathbf{W}_m$ is used as the initial state for running the algorithm when the $(m+1)^{th}$ frame is received.

In order to adopt the presented incremental NMF algorithm to the background modeling problem, role of α should be examined in detail. Let *m* be the number of surveillance frames that used for constructing the initial background representation. Consequently F_m becomes the cost function corresponding to the *m* background frames and f_{m+k} denotes the reconstruction error of the $(m+k)^{th}$ frame. Following this notation, generalization of Eq.(4) for *m*+*k* frames is straightforward and yields Eq.(6):

$$F_{m+k} = (1-\alpha)^k F_m + \alpha(1-\alpha)^{k-1} f_{m+1} + \alpha(1-\alpha)^{k-2} f_{m+2} + \ldots + \alpha f_{m+k}. \quad (6)$$

Note that α controls algorithm's adaptability to content changes. Because α is selected in the interval $(0,1)$, it is straightforward to rank the weights of each frame on the background representation by Eq.(7)

$$\alpha(1-\alpha)^{k-1} < \alpha(1-\alpha)^{k-2} < \ldots < \alpha, \quad (7)$$

where $\alpha(1-\alpha)^{k-i}, i = 1,2,\ldots,k$ denotes the weighting factor of $(m+i)^{th}$ frame.

It should be emphasized that when the number of observed frames, *k*, increases, effect of the initial background model on the new representation decreases. Furthermore, effect of the earlier frames on the representation is smaller than the latest frames, resulting in an adaptive background modeling. We can control adaptation rate of the model to dynamic changes by α. For bigger α, the influence of the last observation on the factorization will be higher.

4 Statistical Background Modeling by Incremental PCA

Principal Component Analysis (PCA) is a method often used to build a low-dimensional representation space spanned by a set of orthogonal vectors. The conventional methods of PCA operate in batch mode. The incremental PCA (IPCA) algorithm extends the static version of PCA modeling to a dynamic and adaptive method by introducing an incremental updating scheme. In this work, the IPCA algorithm proposed in [6] is implemented for dynamic background modeling. Test results obtained by the IPCA and by the proposed INMF are evaluated for comparison purposes.

Let **C** be the $n \times n$ covariance matrix of the data where *n* is the number of frames used for background modeling. It is shown that equality described in Eq.(8) is hold when a matrix **W** contains the eigenvectors of **C** as its columns, and $\mathbf{\Lambda}$ is a diagonal matrix of eigenvalues.

$$\mathbf{CW} = \mathbf{W\Lambda}. \quad (8)$$

Conventionally, the eigenvectors corresponding to the highest eigenvalues, thus *r* columns of **W** where *r*<*n* are used in the PCA representation of a dynamic background.

After construction of the background model, a new data vector $\mathbf{v}'$ can be projected as $\mathbf{h} = \mathbf{W}^T(\mathbf{v}' - \boldsymbol{\mu})$, where μ is the mean vector. The foreground objects are represented by the reconstruction error, $F = |\mathbf{v}' - \mathbf{W}\mathbf{h} + \mu|$.

For dynamic background modeling by IPCA, the impact of the new image must be added to the current model by using an appropriate updating rule. When a new observation vector $\mathbf{v}'$ is received, the mean PCA vector can be updated as in Eq.(9).

$$\boldsymbol{\mu}^{\mathbf{new}} = \alpha\boldsymbol{\mu} + (1-\alpha)\mathbf{v}' = \boldsymbol{\mu} + (1-\alpha)\mathbf{v} . \tag{9}$$

where α and $1-\alpha$ are the updating weights that determines contribution of the previous and new observations to the background representation, respectively. As it is shown in Eq.(9), where $\mathbf{v}$ denotes the new mean-normalized observation data vector, the effects of the old frames on the representation decay exponentially over time. Selection of the parameter α is application-dependent and has to be decided experimentally.

Consequently, the new covariance matrix $\mathbf{C}^{new} = \mathbf{A}\mathbf{A}^T$ can be formed by $r+1$ observation vectors where the matrix **A** and its entities are described by Eq.(10). Note that all of the entities except y_{r+1} are approximated from the eigenvectors of the current model.

$$\mathbf{A} = [y_1, ..., y_{r+1}], \qquad y_i = \sqrt{\alpha\lambda_i}\, w_i , \qquad y_{r+1} = \sqrt{1-\alpha}\,\mathbf{v} \quad i = 1 \ldots r . \tag{10}$$

Eigenvectors and eigenvalues of the background model are updated by eigen-decomposition of the new covariance matrix. Instead of $n \times n$ matrix $\mathbf{C}^{new}$, using $(r+1)\times(r+1)$ matrix $\mathbf{B} = \mathbf{A}^T\mathbf{A}$ for eigen-decomposition problem and then multiplying both sides by **A** leads to Eq. (11).

$$\mathbf{A}\mathbf{A}^{\mathbf{T}}\mathbf{A}\mathbf{e}_{\mathbf{i}} = \mathbf{A}\lambda_i^{new}\mathbf{e}_{\mathbf{i}} . \tag{11}$$

By defining $\mathbf{w}_{\mathbf{i}}^{\mathbf{new}} = \mathbf{A}\mathbf{e}_{\mathbf{i}}$, eigen-decomposition of $\mathbf{C}^{new}$, which requires calculation of new eigenvectors $\mathbf{w}_{\mathbf{i}}^{\mathbf{new}}$ and new eigenvalues λ_i^{new} of the model, can be completed.

5 Test Results

In order to compare the performances of batch-mode NMF, INMF and incremental PCA on dynamic background modeling in surveillance type of video, several tests are carried out on the surveillance video sequences taken from PET2001 database [8].

First test is performed to evaluate the effect of α on the INMF's background representation. Figure 1(a) illustrates distribution of the reconstruction error of each frame, f_m, versus frame number for two different α values. The incremental nature of the algorithm, which makes the effects of the previous frames decay exponentially, allows the choice of a small rank to represent the background adequately. Thus, rank of the representation is set to r=2. The sequence used in this test contains the frames from 800 to 1500 of dataset1 training camera1 sequence of PETS2001 database. First

10 frames are used for the background representation. Small f_m values until frame no 900 shown in Figure 1(a) illustrate that the initial background representation is successful. Moreover, when a motion is detected in the scene, the plot starts to fluctuate. The significant increases in the plot correspond to appearance of a new foreground object whereas the sharp drops refer to the stopping objects that integrated into the background representation. As it is expected, when the contribution of the last observation is small means for smaller values of α , the reconstruction error reaches to higher levels without a major change in the characteristics of the distribution (Figure 1(a)). It should also be noted that, since consecutive frames are very similar in surveillance video, convergence is quickly achieved in a small iteration number.

We have tested the dynamic background modeling performances of the batch NMF and the proposed incremental NMF representations on the same video frames. Rank is set to r=2 for both models and α is set to 0.2 for INMF. Figure 1(b) illustrates distribution of the reconstruction error versus frame number. It is observable that the reconstruction error of INMF remains much smaller than that of batch NMF. Furthermore, although the error is small for both decompositions at the beginning, it never drops to the initial value for the batch NMF. This is because the batch process cannot adapt the background representation according to content changes properly. This makes it unsuitable to an on-line video content analysis. However, the proposed incremental NMF is capable of updating the initial background model according to the dynamic changes. For this reason, the reconstruction error drops to the initial value whenever all of the moving objects become part of the background.

Figure 2 visually demonstrates performance in tracking the foreground objects and updating the background model for the NMF, INMF and IPCA. Figure 2(a) illustrates frame 971 taken from the dataset1 training camera 1 video sequence of PETS2001 dataset. In this frame, a new car (car 1) enters to the scene as the green car (car 2) in the corner starts to leave from the parking lot. In Figure 2b, which corresponds to frame number 1436, car 1 parked to a slot next to the red car and stopped. Car 2 left out its parking slot and moved to the left, thus it is about leaving the scene. In addition, two new walking men exist in this frame. Therefore, it is expected that a powerful method should detect the moving objects which are the two men and car 2 in this scene. In fact, as it is shown in Figure 2(c), (d) and (e), three of the methods are capable of detecting these foreground objects. However, as it is observable in Figure 2(c), the batch NMF also detects the parked car 1 as a foreground object. Furthermore, the old location of car 2 is also not cleared. The reason for their existence is that the batch NMF is not capable of updating the background model and fails to include car 2 into the background model and to remove car 2's old location from the background. However, adaptive updating of the background model is achieved by the INMF successfully (Figure 2(d)). Hence, the proposed INMF is capable of controlling the algorithm's adaptability to dynamic content changes. As it can be seen in Figure 2(e), the IPCA shows a similar performance.

Superiority of the INMF on IPCA becomes much clearer under the illumination changes that may frequently occur in an outdoor surveillance video scene. The distribution of f_m for the frames 2600 to 2990 of dataset2 training camera 2 of PETS2001 database is plotted in Figure 3. Plots are obtained by the INMF with r=2, α =0.05 and IPCA with r=2, α =0.95. As it is shown, the minimum reconstruction error obtained by the IPCA remains much higher than the error of INMF. Furthermore

it makes peaks when the illumination changes significantly. However the minimum reconstruction error achieved by the INMF remains stable within the same video clip. Weakness of the IPCA and robustness of the INMF are visually observable from the Figure 4. Figure 4(a) and (b) illustrate the original video frames 2635 and 2874, respectively. Illumination difference between these frames is recognizable. Figures 4(c) and (d) show the reconstructed difference image obtained by the IPCA with r=2, α =0.95 and obtained by the INMF with r=2, α =0.05, respectively. INMF's ability to remodel the background by adapting it to the illumination changes avoids the appearance of the noise components (Figure 4(d)) that are clearly visible in the scaled difference image for IPCA representation (Figure 4(c)). The main reason behind why IPCA fails in adopting the background to the illumination changes is theoretically IPCA modeling assumes the transformed frames constitute a Gaussian cluster and as it is given by Eq.(10), mean vector of the Gaussian is updated at each iteration. However, illumination changes significantly move the mean vector that can not be incrementally compensated by the IPCA.

6 Conclusions

In this paper, a new approach for dynamic background modeling problem which is based on non-negative matrix factorization is proposed. The proposed representation allows modeling the background successfully and adapting the dynamic scene changes into the background model properly.

Comparison between the conventional batch NMF, the proposed incremental NMF and the incremental PCA representation has been made in order to demonstrate the INMF's success in video surveillance applications. It is concluded that the INMF is much more robust to illumination changes than the IPCA. Test results demonstrate that the INMF is capable of adapting to dynamic background changes within around 0.5 seconds. Currently we are working on deriving new functions in order to decrease the adaptation delay to the order of milliseconds.

References

1. Gutchess, D., Trajkovic, M., Cohen-Solal, E., Lyons, D., Jain, A.: A Background Model Initialization Algorithm for Video Surveillance. Proceedings of International Conference on Computer Vision. (2001)
2. Lee, D.D., Seung, H.S.: Learning the Parts of Objects by Nonnegative Matrix Factorization. Nature, Vol. 401. (1999) 788-791
3. Pascual-Montano, A., Carazo, J.M., Kochi, K., Lehmann, D., Pascual-Marqui, R.D.: Nonsmooth Nonnegative Matrix Factorization. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 42. (2006) 403-415
4. Hoyer, P.O.: Non-negative Matrix Factorization with Sparseness Constraints. Journal of Machine Learning Research, Vol. 5. (2004) 1457-1469
5. Bucak, S.S., Gunsel, B.: Video Content Representation by Incremental Non-negative Matrix Factorization. Submitted to IEEE International Conference on Image Processing. (2007)

6. Li, Y., Xu, J., Morphett, L., Jacobs, R.: An Integrated Algorithm of Incremental and Robust PCA. Proceedings of IEEE International Conference on Image Processing. (2003)
7. Oliver, N., Rosario, B., Pentland, A.: A Bayesian Computer Vision System for Modeling Human Interactions. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22, no.8. (2002) 831-841
8. PET2001 Surveillance Video Database (http://ftp.pets.rdg.ac.uk/)

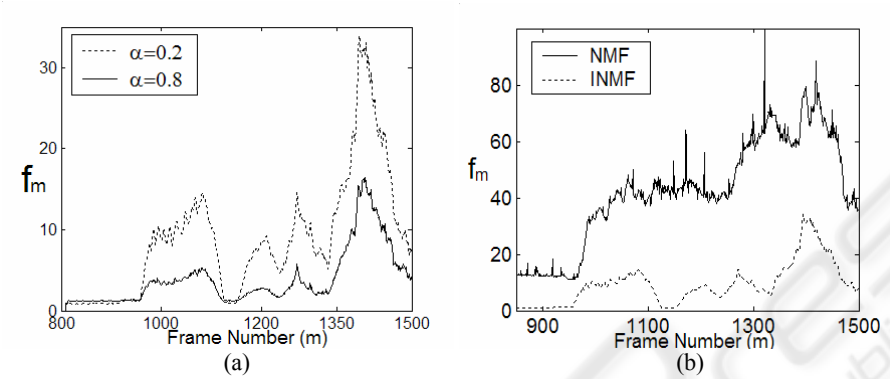


Fig. 1. (a) Distribution of f_m for different α values in INMF representation versus frame number (r=2, frames from 800 to 1500). (b) Distribution of f_m versus frame number for INMF with r=2, α =0.2 and for batch-NMF with r = 2 (frames from 800 to 1500).

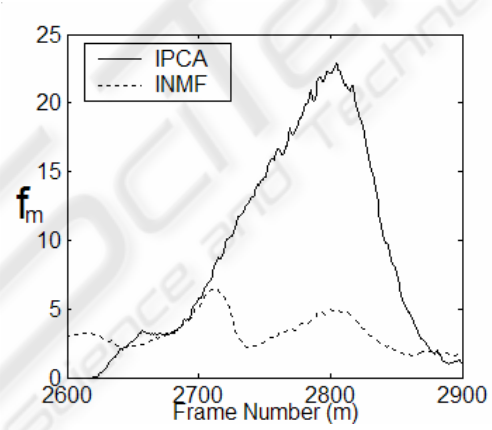
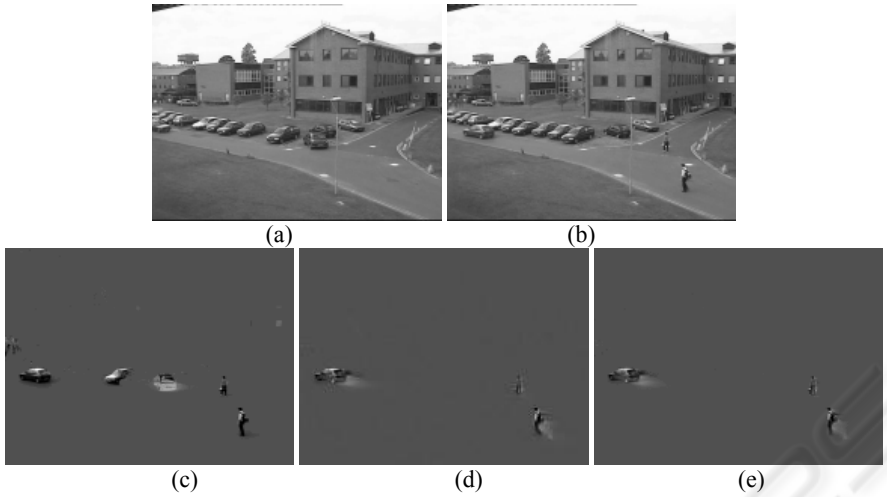


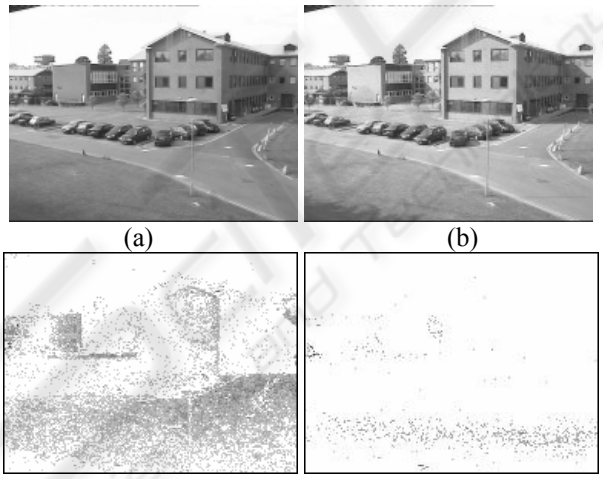
Fig. 3. Distribution of f_m with respect to frame number (frames from 2600 to 2900). The INMF with r=2, α =0.05 and IPCA with r=2, α =0.95 are used for comparison of robustness to illumination changes.



(a) (b)

(c) (d) (e)

Fig. 2 (a) Original video frame 971. **(b)** Original video frame 1436. Reconstructed difference image obtained for the frame 1436 **(c)** by batch NMF with r=2, **(d)** by INMF with r=2, α =0.2 and **(e)** by IPCA with space size r=2, α =0.8.



(a) (b)

(c) (d)

Fig. 4. Robustness to illumination changes: **(a)** Original video frame 2635. **(b)** Original video frame 2874. **(c)** Reconstructed difference image obtained by IPCA with r=2, α =0.95 for the video frame shown in (b). **(d)** Reconstructed difference image obtained by INMF with r=2, α =0.05 for the video frame shown in (b).