Pedestrian Detection Using Global-Local Motion Patterns

Dhiraj Goel and Tsuhan Chen

Department of Electrical and Computer Engineering
Carnegie Mellon University, U.S.A.
dhiraj@cmu.edu, tsuhan@cmu.edu

Abstract. We propose a novel learning strategy called Global-Local Motion Pattern Classification (GLMPC) to localize pedestrian-like motion patterns in videos. Instead of modeling such patterns as a single class that alone can lead to high intra-class variability, three meaningful partitions are considered - left, right and frontal motion. An AdaBoost classifier based on the most discriminative eigenflow weak classifiers is learnt for each of these subsets separately. Furthermore, a linear three-class SVM classifier is trained to estimate the global motion direction. To detect pedestrians in a given image sequence, the candidate optical flow sub-windows are tested by estimating the global motion direction followed by feeding to the matched AdaBoost classifier. The comparison with two baseline algorithms including the degenerate case of a single motion class shows an improvement of 37% in false positive rate.

1 Introduction

Pedestrian detection is a popular research problem in the field of computer vision. It finds its applications in surveillance, fast automatic video browsing for pedestrians, activity monitoring etc. The problem to localize pedestrians in image sequences, however, is extremely challenging owing to the variations in pose, articulation and clothing. The resulting high intra-class variability for the class of pedestrians is further exaggerated by the background clutter and the presence of pedestrian-like upright objects in the scene like trees and windows.

Traditionally, appearance and shape cues have been the popular discernible features to detect pedestrians in a single image. Oren et al. [1] devised one of the first appearance based algorithms using wavelet response, while more recently, histogram of oriented gradients [2] have been used to learn a shape-based model to segment out humans. However, in an uncontrolled environment the appearance cues alone aren't faithful enough for reliable detection.

Recently, motion cues have been gaining a lot of interest for pedestrian detection. In general, pedestrians need to be detected in videos where high correlation between consecutive frames can be used to good effect. While human appearances can be deceptive in a single image, their motion patterns are significantly different from other kinds of motions like vehicles (Fig. 2). The articulation of the human body while in motion due to the movement of limbs and torso can

Y. Yagi et al. (Eds.): ACCV 2007, Part I, LNCS 4843, pp. 220–229, 2007.
© Springer-Verlag Berlin Heidelberg 2007

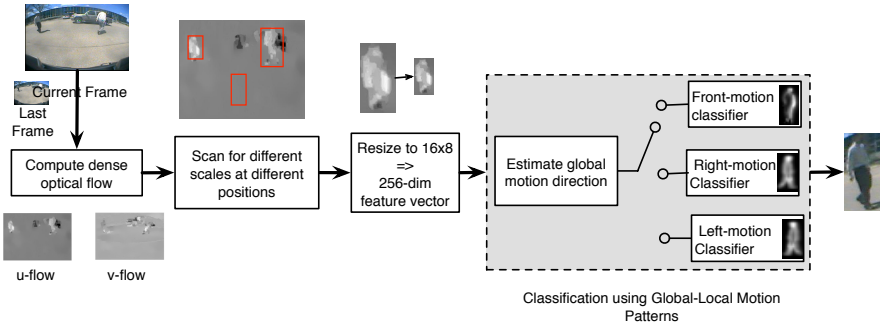


Fig. 1. Overview of the proposed system

provide useful cues to localize moving pedestrians, especially in a stationary cluttered background. To model such a phenomenon, spatio-temporal filters based on shifted frame difference were used by Viola et al. [3], thus, combining the advantages of both shape and motion cues. Fablet and Black [4] used dense optical flow to learn a generative human-motion model while a discriminative model based on Support Vector Machines was trained by Hedvig [5].

The common feature in all the above techniques is that they consider pedestrians as a single class. Though at one hand using human motion patterns circumvents many problems posed by appearance cues, considering all such patterns as a single class can still lead to a very challenging classification problem. In this paper, we present a novel learning strategy to partition the human motion patterns into natural subsets with lesser variability. The rest of the paper is organized as follows: Sect. 2 provides an overview of the proposed method, Sect. 3 introduces the learning strategy based on partitioning the human motion pattern space, Sect. 4 reports the comparison with two baseline algorithms and detection results, and Sect. 5 concludes with a discussion.

2 Overview

Figure 1 gives an overview of the proposed system to detect pedestrian-like motion patterns in the image sequences. Figure 2 illustrates some of the examples of such patterns. Due to high intra-class variability of the flow patterns generated by the pedestrians, modeling all such patterns using a single classifier is difficult. Hence, these are divided into meaningful subsets according to the global motion direction - left, right and frontal. As a result, the classification is divided into two stages. A linear three-class Support Vector Machines (SVM) classifier is trained to estimate the global motion direction. Next, a cascade of AdaBoost classifiers with the most discriminative eigenflow vectors is learnt for each of the global motion subsets. The motion patterns in the same partition share some similarity and hence, intra-class variability for each of these subsets is be less as compared to the whole set, rendering the classification less challenging.

(a) (b)

Fig. 2. (a) Pedestrian sample images along with their horizontal optical flow for right, left and frontal motion subsets. (b) Sample labeled images from the non-pedestrian data and examples of non-pedestrian horizontal flow.

At the time of testing, the dense optical flow image is searched for pedestrian-like motion patterns using sub-windows of different sizes. For every candidate sub-window, first the global motion direction is estimated using the linear three-class SVM classifier. Thereafter, it is tested against the matching AdaBooost classifier.

2.1 Computing Dense Optical Flow

Dense optical flow is used as a measure to estimate motion between consecutive frames. Though numerous methods exist in the literature to compute dense flow, 2-D Combined Local Global method [8] was chosen since it has been shown to provide very accurate flow. Furthermore, using bidirectional multi-grid strategy, it can work in real-time [9] at upto 40 fps for 200x200 pixels image. The final implementation used for pedestrian detection incorporates a slight modification in the weighting function of the regularization term as mentioned in [6].

2.2 Training Data

The anatomy of the learning algorithm necessitates a pedestrian data set labeled according to the global motion. For this purpose, the CASIA Gait database [7] was chosen. A total of eight global motion directions were considered that were merged to give three dominant motions - left, right and frontal (Fig. 2(a)). The left and the right motion subset capture the lateral motion while the motion perpendicular to the camera plane is contained in the frontal motion subset.

Dense optical flow was computed for the videos and the horizontal, u, and the vertical, v, flows for the labeled pedestrians were cropped. The collection of these flow patterns formed the training-test data for the classification. Specifically, the frontal motion subset had 2500 training data samples and 1000 test data samples. The other two motion subsets had 4800 training data samples and 2000 test data samples each. The cropped data samples were resized to 16x8 pixels, normalized to lie in the range $[-1, 1]$ and concatenated to form a 256 dimension feature vector - $[u_1, u_2, \ldots, u_{128}, v_1, v_2, \ldots, v_{128}]$.

The non-pedestrian data was generated by hand-labeling sub-windows with non-zero flow in the videos containing moving vehicles. To automate the process,

an Adaboost classifier was trained for the set of all pedestrian and non-pedestrian data and was run on other videos to generate additional non-pedestrian flow patterns (from the false positives). The non-pedestrian data samples are resized and normalized in the same way as the pedestrian data. Approx. 120,000 such samples were generated, with some examples shown in Fig. 2(b).

3 Classification Strategy

This section describes the classification strategy to distinguish the motion patterns of pedestrians from other kinds of motions like that of vehicles etc. As illustrated in Fig. 1, it is divided into two stages - estimating the global motion direction (Section 3.1) followed by testing against the discriminative classifier (Section 3.2). Training procedure for the latter has been described in [6].

The final detection performance depends on the accuracy of both the stages and is greatly influenced by the taxonomy of the pedestrian motion patterns. A maximum of eight possible motion classes were considered as shown in the Fig. 2(a). Building a discriminative classifier for each of them results in a group of classifiers that are highly discriminative for the motion direction they are trained for. Thus, the accuracy in estimating the motion direction becomes crucial to the overall performance, i.e. the sub-window containing strictly left moving pedestrian should be fed to the classifier trained to detect strictly left moving pedestrians. However, it is very difficult to reliably estimate the motion direction in these eight subsets. Thus, the detection rate of the classifier as a whole degrades. The natural modification is to merge the different motion subsets such that the motion direction can be estimated faithfully but at the same time intra-class variability is kept low. Splitting the motion patterns into three subsets - left, right and frontal - gave the best performance.

3.1 Estimating Global Motion

In order to decide which motion-specific discriminative classifier to use, it is important to first estimate the global motion. The mean motion direction for the pedestrian data was found to be unreliable in achieving such an objective. Hence, a linear three-class SVM classifier was trained. This classifier acts as more of a switch that assigns the queried data samples to their appropriate classifiers that have been specifically trained to handle those particular flow patterns.

The labeled pedestrian data is used to train this switch classifier. The same number of training data samples, about 2000 each, was used for all the three classes to obviate bias towards any particular class. Further, each of the classes themselves contain the same proportion of different motions contained within them. For example, the left class contains the same number of samples for strict left motion, left front at 45^o and left back at 45^o. Figure 3 shows the class confusion matrix for the learned model. 348 support vectors were chosen by the model that is less than 6% of the number of training data samples, indicating a well generalized classifier.

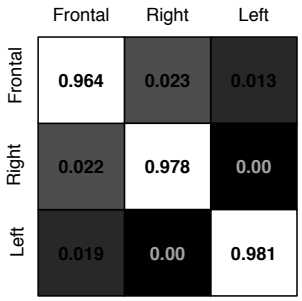


Fig. 3. Class confusion matrix for estimating the global motion direction using the three-class linear SVM classifier

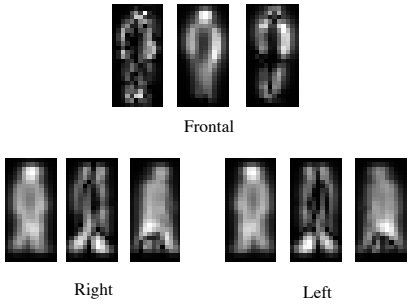


Fig. 4. Magnitude of the mean and the first two eigenflow vectors of the horizontal optical flow for the training pedestrian data

The trained switch classifier is used to allocate non-pedestrian data for each of the motion classes for training the discriminative motion-specific classifiers. Out of 120,000 data samples, about 75,000 got classified as belonging to the frontal motion, 25,000 were categorized as left motion class while the remaining 20,000 as having right motion.

3.2 Learning the Discriminative Classifiers

This section describes the learning procedure to train the discriminative motion-specific classifiers. In total, three separate classifiers are learnt, one for each global motion. The learning process is the same for all of them. Hence, for the sake of clarity, motion-specific term has been dropped in this section and whenever pedestrian and non-pedestrian data is mentioned, it refers to the data belonging to a particular global motion, unless stated otherwise. It is worth mentioning that the symmetrical properties of left and right classifiers can be exploited by training the classifier for one and using it's mirror image (after changing the sign for horizontal motion) for the other.

Weak Classifier. Principal Component Analysis was done separately on the pedestrian and non-pedestrian data to obtain the eigenvectors for the optical flow, known as eigenflow [10]. Figure 4 shows the magnitude of the mean and the first two u-flow eigenvectors for each of the three global motions. As is evident, the mean flows represent the global motion while the eigenflow vectors capture the poses and the articulation of the human body, especially the movement of the limbs. For the frontal motion, the mean is not that informative since it contains both front and backward moving pedestrians.

Using all the eigenflow vectors, 256 for each of the pedestrian and non-pedestrian data, we have a total of 512 eigenflow vectors that act as a pool of features for AdaBoost. Taking the magnitude of correlation between the training data x and an eigenflow vector z_j and finding the optimum threshold θ_j that

Table 1. Feature selection and training AdaBoost classifier

- Given the training data $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ where x_i is the eigenflow and y_i is 0 for non-pedestrian and 1 for pedestrian examples.
- Initialize the weights $w_{1,i} = \frac{1}{2l}, \frac{1}{2m}$ for $y_i = 0, 1$ respectively, where l and m are the number of pedestrian and non-pedestrian examples.
- for $t = 1, \ldots, T$
  1. Normalize the weights $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$
  2. Select the best weak classifier h_t with respect to the weighted error: $\epsilon_t = min_j \sum_i w_i |h_j - y_i|$
  3. Update the weights: $w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$
     where $e_i = 0$ if example x_i is correctly classified by h_t, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The strong classifier is given by:

$$C(x) = \begin{cases} 1, \text{if } \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0, \text{otherwise.} \end{cases} \tag{2}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

minimizes the overall classification error would yield a weak classifier h_j.

$$h_j(x) = \begin{cases} 1, \text{if } |x^T z_j| \lessgtr \theta_j \\ 0, \text{otherwise.} \end{cases} \tag{1}$$

Feature Selection and AdaBoost. The procedure to choose the most discriminative of the weak classifiers, as illustrated in Fig. 5(a) is motivated by the face detection algorithm proposed in [11]. Table 1 describes the complete algorithm. The final strong classifier is a weighted vote of the weak classifiers (Eq. (2)).

Figures 5(b), (c) and (d) depict the horizontal component of the two eigenflow features selected by this algorithm for each of the global motion subset. The selection of the most discriminative vectors follows a similar trend in all the three cases. While the first one responds to motion near the boundary, the second one captures the motion within the window. It is also interesting to note the pattern at the bottom of the first eigenflow vectors - those belonging to the right and left subsets take into account the spread of the legs in the lateral motion while the one for the frontal motion restricts any such articulation. Individually, they may perform poorly but as a combination, they can perform much better.

Table 2 juxtaposes the false positive rate (FPR) of the GLMPC classifier with two other classifiers for a fixed detection rate of 98%. The first one is the linear SVM classifier that is clearly outperformed in both speed and accuracy. 13,313 support vectors were chosen by the linear SVM that is more than 50% of the training data, an indication of a poorly generalized classifier. Besides, such a

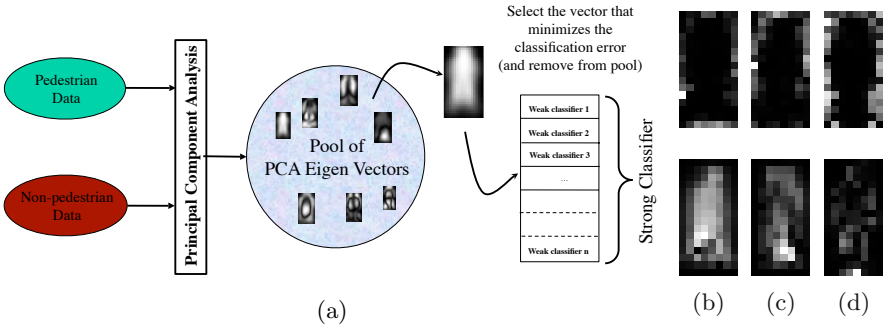


Fig. 5. (a) Feature Selection using AdaBoost. (b), (c) and (d) Two u-eigenflow vectors selected by AdaBoost for the Right, Left and Frontal subsets respectively.

Table 2. False positive rate for the different classifiers for the detection rate of 98%

	SVM	LMPC	GLMPC
False Positives (%)	62.3	1.16	0.74

high number of support vectors would result in about 1.3 million dot products per frame, assuming 100 candidate sub-windows in a frame. On the other hand, classification using GLMPC requires only 348 dot products for the three-class SVM switch and 35 dot products for AdaBoost cascade (full cascade in the worst case). The other classifier considered for comparison is the degenerate case of the proposed algorithm, that we refer as Local Motion Pattern Classifier (LMPC) [6], when all the pedestrian data is considered as one single class. GLMPC provides a reduction of 37% in FPR that is further amplified by the fact that they may hundreds of candidate sub-windows in a frame.

Cascade of AdaBoost Classifiers. In general, in any scene, flow patterns that share no resemblance with human motion should be discarded quickly, while those that share greater similarity require more complex analysis. A cascade of AdaBoost classifiers [11] can achieve this. The early stages in the cascade have a lesser number of weak classifiers and hence, aren't too discriminative but are really fast at classification. The later stages consist of more complex classifiers with larger number of weak classifiers. To be labeled as a detection, a candidate data sample has to pass through all the stages. Hence, the classifier spends most of the time analyzing difficult motion patterns and rejects easy ones quickly.

In our implementation, there are two stages in the cascade for each of the global motion classifiers. The same pedestrian data was used across all stages. For training the classifier, the ratio of pedestrian to non-pedestrian data (for both training and test data) was kept at one for the left and right motion subsets and 0.5 for the frontal motion. Non-pedestrian data for the next stage in the cascade is generated by collecting the false positives after running the existing classifier on different videos taken from both static and moving cameras. The final frontal classifier has 5 weak learners in the first stage and 20 in the second.

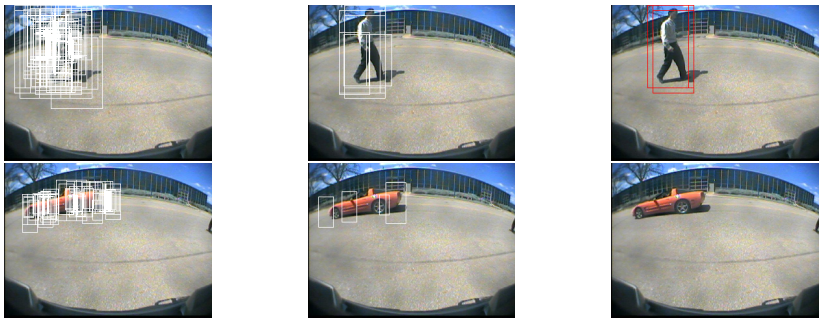
The corresponding numbers for the right and the left motion classifiers are 10 and 25, and 10 and 20 respectively.

4 Experiments

For detecting human motion patterns, the dense optical flow image is searched with sub-windows of different scales, seven in total. Every scale size also has an associated step size. Naturally, larger sub-windows have bigger steps size to prevent redundancy due to excessive overlap between neighboring sub-windows. Knowing a priori, the camera orientation can greatly reduce the search space since the pedestrians need to be looked for only on the ground plane. Exploiting such an information reduced the total number of scanned sub-windows in the image by almost half. Finally, only the candidate sub-windows that satisfy the minimum flow thresholds are resized and normalized, before feeding to the classifier. Again, these thresholds vary with the scale size as larger sub-windows search for near-by pedestrians that should appear to move faster due to parallax.

Figure 6 depicts the detection results by linear SVM, LMPC and GLMPC classifier after the first stage in the cascade. The overlapping windows have not been merged to show the all the detected sub-windows. As is evident, the GLMPC is able to localize the pedestrians much better than any of the two methods and in addition, gives less false positives.

The full cascade GLMPC classifier was tested for pedestrian patterns in different test videos and works at 2fps on a Core 2 Duo 2 GHz PC. Figure 7 shows some of the relevant results. The algorithm was tested with multiple moving pedestrians in the presence of other moving objects, mainly cars and is able to detect humans in different poses and moving at different pace (Fig. 7(a)). The occluding objects can lead to false rejections since the flow in the concerned sub-window doesn't conform to the pedestrian motion. This is evident in the



(a) SVM (b) LMPC (c) GLMPC

Fig. 6. Comparison of the performance of GLMPC classifier with linear SVM and LMPC after Stage 1 in the cascade. Color coding - white if direction is not known, red for right moving pedestrians, yellow for left and black for frontal motion.

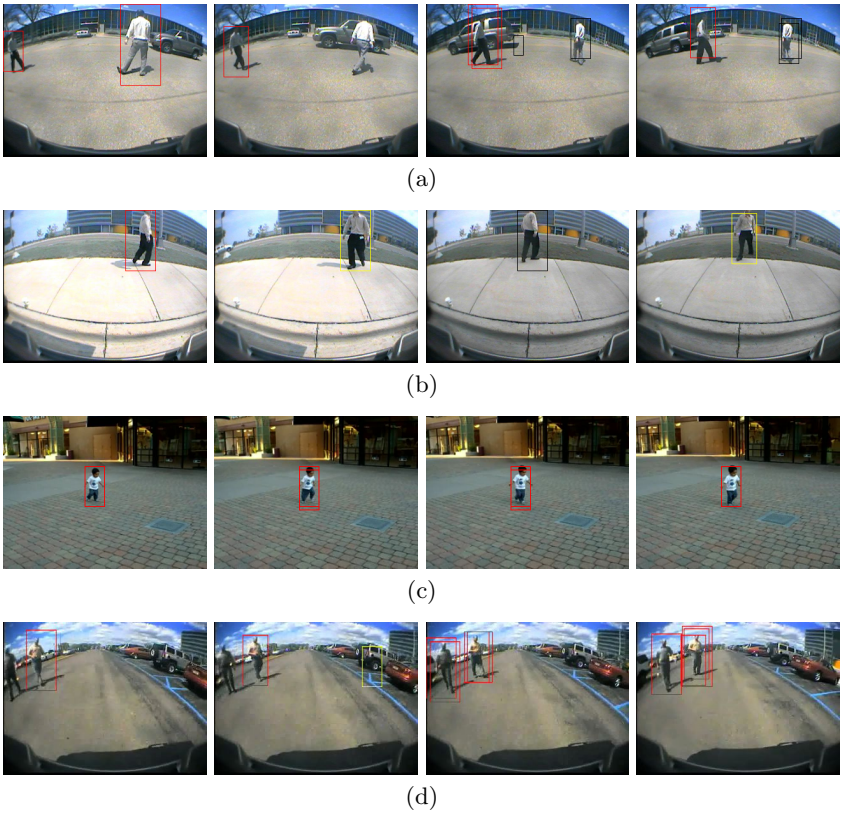


Fig. 7. Final detection results without merging the overlapping detections

second image in Fig. 7(a). Stationary and far-off pedestrians that are moving very slowly can also be missed owing to their negligible optical flow.

The system is also robust to illumination changes (Fig. 7 (b)) and can detect moving children (Fig. 7(c)) even though the training data was composed of only adult pedestrians. Moreover, notice the panning of the camera over time in the image sequence, illustrating the robustness of the system towards small camera motion. The videos captured from a slow moving car were also tested and the system still manages to detect pedestrians (Fig. 7 (d)).

5 Discussion

A novel learning strategy to detect moving pedestrians in videos using motion patterns was introduced in the paper. Instead of considering all human motion patterns as one class, they were split into three meaningful subsets dictated by the global motion direction. A cascade of AdaBoost classifiers with the most discriminative eigenflow vectors were learnt for each of these global motion

subsets. Further, a linear three-class SVM classifier was trained that acts as a switch to decide which Adaboost classifier to choose to determine if a pedestrian is contained in the candidate sub-window.

It was shown that the proposed algorithm is far superior to the linear SVM and provides an improvement of 37% in FPR as compared to LMPC. Moreover, the proposed system has been shown to be robust to slow illumination changes, camera motion and can even detect children. Apart from conspicuous advantages of accuracy, GLMPC allows for extensibility to incorporate new pedestrian motion like jumping without retraining the whole classifier again. Only a couple of changes would be required. The first would be to retrain the motion switch multi-class SVM classifier to take into account the new motion type. The next would be to train a new AdaBoost classifier to discriminate between the jumping motion of the pedestrians and other kinds of motions. The already trained classifiers for left, right and frontal motion can be used in their original form.

An important area of research for the future work would be to compute the ROC curve for the classifiers like GLMPC that don't have a single global threshold. Work on similar lines has been done by Xiaoming et al. [10].

References

1. Oren, M., Papageorgiou, C., Sinha, P., Osuna, E., Poggio, T.: Pedestrian detection using wavelet templates. CVPR, 193–199 (1997)
2. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. CVPR 1, 886–893 (2005)
3. Viola, P., Jones, M., Snow, D.: Detecting Pedestrians Using Patterns of Motion and Appearance. ICCV 2, 734–741 (2003)
4. Fablet, R., Black, M.J.: Automatic Detection and Tracking of Human Motion with a View-Based Representation. ECCV 1, 476–491 (2002)
5. Sidenbladh, H.: Detecting Human Motion with Support Vector Machines. ICPR 2, 188–191 (2004)
6. Goel, D., Chen, T.: Real-time Pedestrian Detection using Eigenflow. In: IEEE International Conference on Image Processing, IEEE Computer Society Press, Los Alamitos (2007)
7. `http://www.cbsr.ia.ac.cn/Databases.htm`
8. Bruhn, A., Weickert, J., Schnörr, C.: Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optic Flow Methods. IJCV 61, 211–231 (2005)
9. Bruhn, A., Weickert, J., Kohlberger, T., Schnörr, C.: A Multigrid Platform for Real-Time Motion Computation with Discontinuity-Preserving Variational Methods. IJCV 69, 257–277 (2006)
10. Liu, X., Chen, T., Kumar, B.V.: Face authentication for multiple subjects using eigenflow. Pattern Recognition 36, 313–328 (2003)
11. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. CVPR (2001)