Computers and Electrical Engineering 40 (2014) 1072–1079

Contents lists available at ScienceDirect

Computers and Electrical Engineering

journal homepage: www.elsevier.com/locate/compeleceng

Pedestrian detection and tracking using temporal differencing and HOG features ☆

Tudor Barbu *

Institute of Computer Science of the Romanian Academy, Iași, Romania

ARTICLE INFO

Article history:
Received 31 August 2012
Received in revised form 5 December 2013
Accepted 9 December 2013
Available online 6 January 2014

ABSTRACT

This article proposes a multiple human detection and tracking approach. A moving person identification technique is provided first. The video objects are detected using a novel temporal differencing based procedure and several mathematical morphology-based operations. Then, our technique determines what moving image objects represent pedestrian people, by testing several conditions related to human bodies and detecting the skin regions from the movie frames. A robust human tracking method using a Histogram of Oriented Gradient (HOG) based template matching process is then introduced in our paper. Some person detection and tracking experiments and method comparisons are also described.

1. Introduction

Person detection and tracking in movie sequences has become a high-interest computer vision field in the last decades [1,2], representing the most important sub-class of the moving object detection and tracking research domain.

Person detection task consists of identification of the human presence in video streams and differentiate humans from non-human video objects. Human tracking locates the instances of each detected person in the frames of the analyzed movie. Detecting persons in images and movie sequences represents a very challenging task, being complicated by various factors, like video camera position, variable people appearance, wide range of poses adopted by human beings, variations in brightness, illumination, contrast levels or backgrounds, and person occlusions [1–3]. Over the last ten years, the task of detecting and tracking people has received a very considerable interest. Some significant research has been devoted to locating and tracking humans in static and dynamic images, since a lot of applications involve humans' locations and movements [1,2].

Numerous person identification and tracking techniques have been developed recently [3]. Human detection has been approached using algorithms based on frame differences [1,4], background subtraction [5], Partial Least Squares Analysis [6], Haar Wavelets with Support Vector Machines (SVM) [7], Histogram of Oriented Gradients (HOGs) based features [8], Hough transforms [9], Point detectors [10] and Active Contours (Snakes) [11]. Various person tracking methods have been also proposed. Most important tracking techniques are based on Mean-Shift algorithms [12], Kalman filtering [13], Hidden Markov Models (HMM) [14], Optical Flow [15], motion tracking [14] and human matching [16]. Human detection and tracking has also a wide variety of computer vision application areas. The most important of them are: video surveillance and security systems, biometrics, law enforcement, human–computer interaction (HCI), video indexing and retrieval, medical imaging, robotics and augmented reality.

Several person detection and recognition techniques are proposed in our past works. We considered some face detection [17], face recognition [18] and human skin identification techniques [17] in those papers. Video object detection and tracking

☆ Reviews processed and approved for publication by Editor-in-Chief Dr. Manu Malek.

* Tel.: +40 0762675099.
E-mail address: tudbar@iit.tuiasi.ro

0045-7906/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compeleceng.2013.12.004

domain was also investigated by us [19]. The results of these approaches are unified in this new computer vision task of human body detection and tracking. The skin features are combined to video motion estimation solutions to produce an efficient person detection and tracking. In this paper we propose an automatic multiple moving human detection and tracking technique for motionless camera video sequences. The computer vision system developed here identifies only the moving people in upright position (pedestrians) and not all persons from the video frames. Because it performs a robust pedestrian detection and tracking in fixed camera videos, our system can be successfully applied to video surveillance.

This article is composed of five main sections, including this introduction, and three other sections containing the acknowledgements, references and author's biography. The next section introduces a multiple moving video object detection approach using a novel temporal differencing algorithm and some mathematical morphological operations, in its first subsection. In the second subsection, each identified image object is classified either as person or non-person, using skin segments and some conditions related to human body. The third section describes a video tracking process performed on the identified pedestrians. Our tracking technique uses a HOG-based moving object matching procedure. In the fourth section some detection and tracking experiments and method comparisons are presented. The fifth section contains the paper conclusions.

2. A multiple human identification method

The considered system has to solve the following computer vision task: to locate and track all the walking human beings from a video sequence recorded with a static camera. First, the movie has to be pre-processed before performing the required video analysis on it. The human tracking needs to be performed within each video shot, so a temporal segmentation is applied first. In our previous works we provided some shot detection techniques that can be used successfully in this case [20]. If an undesired amount of noise is present in the video frames, some denoising operations can be performed. We proposed some PDE-based noise removal and restoration techniques [21], which are quite useful here, facilitating the next object detection process.

The multiple moving person detection procedure consists of two main phases. The first one, representing a foreground segmentation of the filtered video sequence, is described in the next subsection, while the identification of those video objects representing persons is performed in the second one.

2.1. Moving object detection approach

The video frames of the analyzed sequence are converted into the grayscale form, the set $\{Im_1,\ldots,Im_n\}$ being obtained. The video motion of this frame sequence is then estimated using a frame-difference method [1,2,4].

We consider a temporal differencing algorithm. The difference of two consecutive video frames indicates the motion between them, the resulted non-black image zones representing the moving regions. Such a moving region may not represent an entire image object. That happens because both the foreground and the background of the frame can be composed of more homogeneous regions, characterized by various intensities. We note the frame difference as $Fd(i,j) = Im_i - Im_j$, $\forall i,j \in [1,n]$, $i \neq j$. Each moving object that is present in the frames Im_i and Im_j, is represented in $Fd(i,j)$ by some non-black regions. Its high-intensity pixels (having greater values than background) are displayed in $Fd(i,j)$ at the locations occupied in Im_i, while its low-intensity pixels are displayed in $Fd(i,j)$ at their positions in Im_j. The frame difference is then converted into the binary format using a properly chosen threshold value, T:

$$Fd_b(i,j) = \begin{cases} 1, & \text{for } Fd(i,j) \geqslant T \\ 0, & \text{for } Fd(i,j) < T \end{cases}, \forall i,j \in [1,n], i \neq j. \tag{1}$$

A mathematical morphology based process is next applied to image $Fd_b(i,j)$ [22]. The dilation morphological operation is very useful in this case, producing the dilated binary image as follows:

$$Fd_m(i,j) = Fd_b(i,j) \oplus Sq = \bigcup_{s \in Sq} Fd_b(i,j)_s \tag{2}$$

where Sq is a $[k \times k]$ structuring element and the symbol $\oplus$ represents the dilation morphological operator [22]. The connected components of the dilated image $Fd_m(i,j)$ are then determined and those representing errors provided by still remaining noise or undesired camera motion are removed. We consider the following categories of connected components to be discarded: small white spots – connected components whose area is under a low threshold; components whose bounding rectangle has one of the dimensions under a threshold value; components characterized by a low solidity (ratio between the region area and its bbox).

We note the resulted morphologically processed image as $Fd_m^p(i,j)$. The proposed detection algorithm identifies the moving objects from Im_1 to Im_{n-1}, first. Thus, at each step i, our approach identifies the video objects of Im_i, using the next two frames, Im_{i+1} and Im_{i+2}. The corresponding morphologically processed frame difference images $Fd_m^p(i,i+1)$ and $Fd_m^p(i,i+2)$ are computed, their intersection being also computed. The *intersection* of two binary images is defined as the image having the same pixel values where the two images coincide and 0 where they differ:

$$(Fd_m^p(i,i+1) \cap Fd_m^p(i,i+2))[x,y] = \begin{cases} Fd_m^p(i,i+1)[x,y], & \text{for } Fd_m^p(i,i+1)[x,y] = Fd_m^p(i,i+2)[x,y] \\ 0, & \text{for } Fd_m^p(i,i+1)[x,y] \neq Fd_m^p(i,i+2)[x,y] \end{cases}. \tag{3}$$

The connected components of the intersection $Fd_m^p(i,i+1) \cap Fd_m^p(i,i+2)$ correspond to all high-intensity regions of the moving objects of the video frame Im_i. The low-intensity components of its moving objects are determined using a similar procedure. They correspond to the connected components of $Fd_m^p(i+1,i) \cap Fd_m^p(i+2,i)$. Moving objects of Im_i are identified by computing the sum of intersections:

$$Ob(i) = (Fd_m^p(i,i+1) \cap Fd_m^p(i,i+2)) + (Fd_m^p(i+1,i) \cap Fd_m^p(i+2,i)). \tag{4}$$

The connected components of the binary image $Ob(i)$ correspond to the moving objects of frame Im_i. The last step of the detection task consists of the localization of those objects in Im_n. We apply a backward identification process that is modeled as:

$$Ob(n) = (Fd_m^p(n,n-1) \cap Fd_m^p(n,n-2)) + (Fd_m^p(n-1,n) \cap Fd_m^p(n-2,n)). \tag{5}$$

Each binary image $Ob(i)$ contains the same number of connected components. The bounding box of each one is then identified. The sub-images of frame Im_i corresponding to these bounding rectangles represent its moving objects (foreground). Such a foreground detection example is described in Fig. 1.

2.2. Video analysis of the detected objects

Now, we have to decide which of the detected video objects represent pedestrians. We consider a set of conditions related to human body which must be satisfied by each object representing a walking person. The first condition is that the height of the bounding box of the image object representing a human being must be at least two times greater than its width. The second condition is that the object solidity has to exceed 50%. Let $\{ob_1,\ldots,ob_{n_i}\}$ be the objects corresponding to $Ob(i)$, and $\{I(ob_1),\ldots,I(ob_{n_i})\}$ the set of their sub-images. So, the above conditions could be formalized as follows:

$$\begin{cases} h(ob_j) \geqslant 2 \cdot w(ob_j) \\ h(ob_j) \cdot w(ob_j) \leqslant 2 \cdot Area(ob_j) \end{cases}, \forall j \in [1,n_i] \tag{6}$$

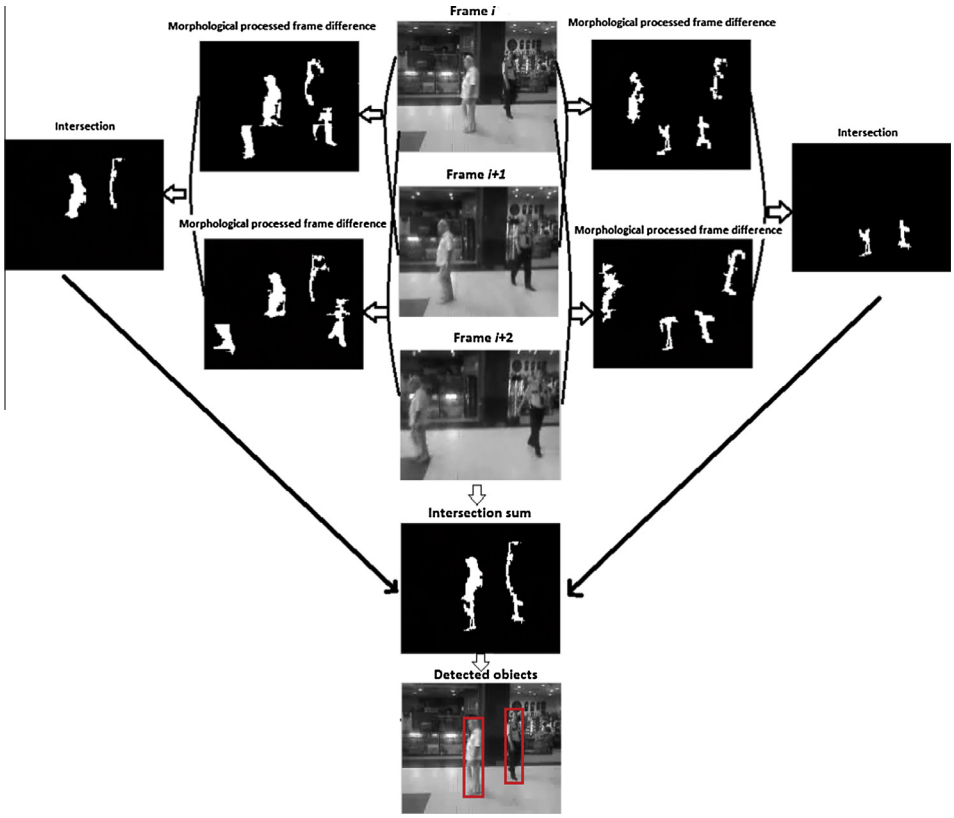

Fig. 1. Moving object detection example.

where $h(ob_j)$ and $w(ob_j)$ are the height and the width of image $I(ob_j)$. Let $Obj(i) \subseteq Ob(i)$ the binary image containing the connected components satisfying (6) only.

The third condition is related to the presence of skin regions. Any video object must contain skin regions to be considered a human body. So, we use a skin detection technique for color images. In the last decades, extensive research has focused on human skin detection, representing the process of identifying skin-colored regions in a static image or frame. The most popular skin recognition technique was proposed by Fleck and Forsyth in 1996 and it is based on explicitly defined regions [23]. We also provided novel skin identification solutions based on decision rules in our previous works [17]. Our detection technique defines explicitly the human skin regions, applying certain restrictions to the channels of the color space, each component of a skin pixel being restricted to a specific interval [17].

Despite representing one of the most used color spaces for processing and storing of digital image data, RGB is not a favorable choice for skin color analysis, because of the high correlation of its channels and the mixing of luminance and chrominance data. For this reason our skin segmentation technique uses HSV and YC_rC_b color spaces. The RGB image is first converted into the Hue Saturation Value format, by computing the three components using the well-known conversion equations. The components, H, S and V, are obtained as three matrices whose coefficients belong to $[0,1]$ interval. Next, a conversion in YC_rC_b color space is performed. In fact, it does not represent an absolute color space, but a way of encoding the RGB information. In this format Y represents the luminance, C_r is the blue-difference and C_b is the red-difference chroma component. These components of the color space are computed as linear combinations of R, G and B channels. The computation formulas of the chroma components, C_r and C_b, have the general form $\alpha \cdot R + \beta \cdot G + \gamma \cdot B + 128$, where coefficients α, β, $\gamma \in [-0.5, 0.5]$. We select empirically some proper values for these coefficients and get the following components:

$$\begin{cases} C_r = 0.16 \cdot R - 0.25 \cdot G + 0.45 \cdot B + 128 \\ C_b = 0.44 \cdot R - 0.34 \cdot G - 0.08 \cdot B + 128 \end{cases}. \tag{7}$$

Next, a set of restrictions is applied on these two components and the hue. Each pixel of the analyzed $[X \times Y]$ image belongs to a human skin segment if the corresponding values in C_r, C_b and H are situated in some intervals. The following binary image, whose white regions correspond to skin segments, results:

$$Sk(i,j) = \begin{cases} 1, & C_r(i,j) \in [148, 167], \quad C_b(i,j) \in [146, 189], H(i,j) \in [0.025, 0.1] \\ 0, & otherwise \end{cases}, i \in [1, X], j \in [1, Y]. \tag{8}$$

The connected components of Sk may represent some human body parts, so the detection of skin regions indicates the human presence. Obviously, this skin recognition method works on the colored versions of the video frames Im_i. If these skin regions are identified in the moving object locations, then those video objects represent persons. Let those detected human objects of $Obj(i)$ be noted as $\{H_1^i, \ldots, H_K^i\}$. Their corresponding sub-images are $\{I(H_1^i), \ldots, I(H_K^i)\}$, where $K \leqslant n_i$ value represents the number of human beings on each frame.

3. Pedestrian tracking based on HOG features and human matching

In this section we propose a video tracking technique for the previously detected moving persons. For each video frame one knows the moving objects representing people, but the instances of each human in the successive frames are still unknown. The considered tracking method determines for any pedestrian from a given frame its instance in the next video frame.

We consider a HOG-based template matching technique for video tracking. For each frame Im_i we have obtained the sequence of pedestrians $\{H_1^i, \ldots, H_K^i\}$. Now, for each $j \in [1, K]$, we must find in the consecutive frame Im_{i+1} the human video object $H_l^{i+1} \approx H_j^i$, where $l \in [1, K]$ and the symbol $\approx$ represents the image object content similarity. Let us note $l = ind_i(j)$. The index value $ind_i(j)$ must be determined for all i and j values. Our human matching technique compares the image $I(H_j^i)$ with all the sub-images $I(H_l^{i+1})$ from the successive frame and identifies the *closest* one.

There are various solutions to this video object matching task. Most of them represent pixel differencing based techniques. Thus, one could compute *SAD* (sum of absolute differences), *MAD* (mean absolute distance), *MSD* (mean squared distance), and *NCC* (normalized cross-correlation) [17,24] between these sub-images. The problem that arises here is that the sub-images of the same person in different video frames could have different sizes. Because of its moving, both the height and the width of a walking person may vary from frame to frame. Thus, taking bigger steps, moving arms away from the body or moving toward camera represent factors which increase the area of sub-image $I(\cdot)$ of that pedestrian.

This means that the above mentioned approaches, using pixel by pixel correspondences, are not well-suited for our human body matching task, because they can compare same-size images only. The images $I(H_j^i)$ have to be resized each time a SAD, MAD or another operation of this kind is applied to them. Image resizing implies a loss of information and could increase the matching error rate.

Therefore, a better solution in this case is to compute some fixed-size feature vectors of these images. Thus, histogram-based image feature vectors are better suited for this case. Unfortunately, some histograms, such as the popular color (grayscale) histograms, do not represent robust image content descriptors. Two different persons could have the same histogram. For this reason we consider using *Histograms of Oriented Gradients* (*HOG*) for human object featuring [8]. Histogram

of Oriented Gradients represents a robust feature descriptor used in computer vision for the purpose of object detection. Numerous HOG-based pedestrian detection techniques have been developed in recent years [8,25]. Unlike these techniques, our approach uses HOG in the human tracking stage, not in the detection stage.

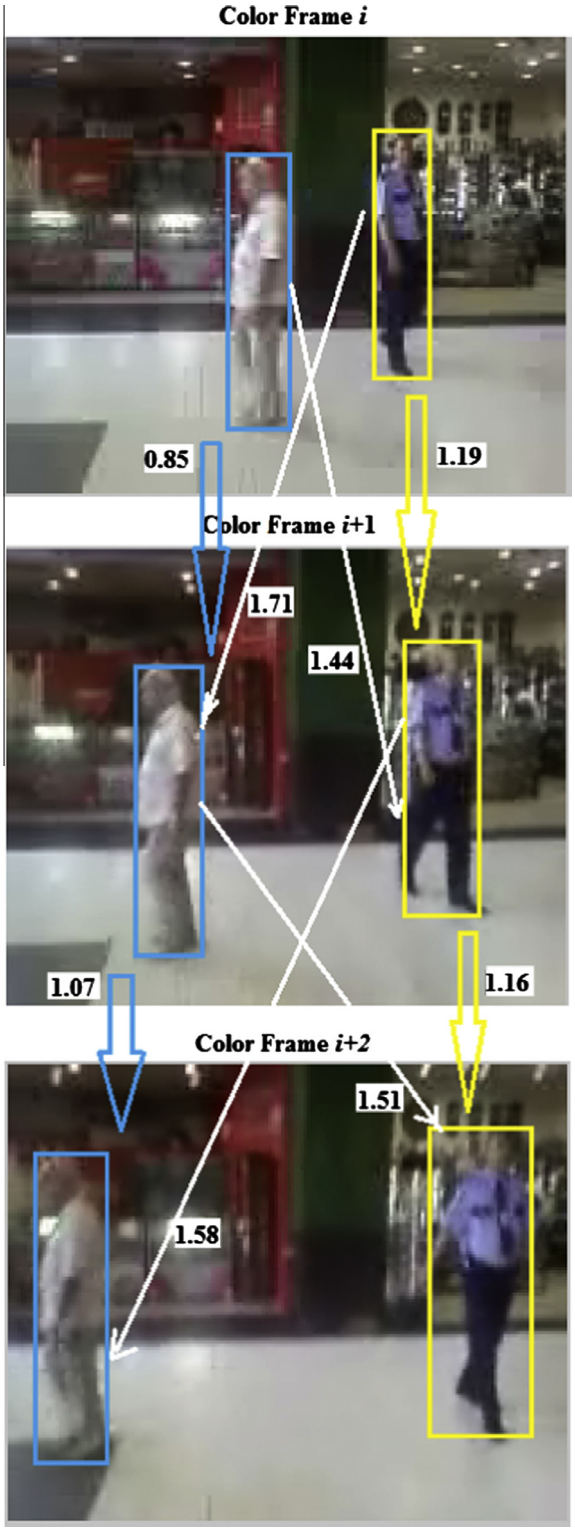

Fig. 2. Pedestrian tracking example.

For each image $I(H_j^i)$, a HOG-based feature vector is computed. First, the image gradient values, representing directional changes in the grayscale intensity in the image, are computed. The gradient vector is formed by combining the partial derivatives of the image in the x and y directions:

$$\nabla I(H_j^i) = \left(\frac{\partial I(H_j^i)}{\partial x}, \frac{\partial I(H_j^i)}{\partial y} \right), \forall i \in [1, n], j \in [1, K] \tag{9}$$

where the gradients in x and y directions can be computed by applying the $1D$ centered, point discrete derivative mask in the horizontal and vertical directions:

$$\begin{cases} \frac{\partial I(H_j^i)}{\partial x} = I(H_j^i) * [-1 \quad 0 \quad 1] \\ \frac{\partial I(H_j^i)}{\partial y} = I(H_j^i) * [-1 \quad 0 \quad 1]^T \end{cases}. \tag{10}$$

The gradient orientations of the image are computed as $\theta = \arctan\left(\frac{\partial I(H_j^i)}{\partial x}, \frac{\partial I(H_j^i)}{\partial y} \right)$. The image $I(H_j^i)$ is then divided into cells [8]. For each cell, a local $1D$ histogram of gradient directions (orientations) over the pixels of the cell is computed. We consider 9 bins for the local histogram. The histogram channels are evenly spread over 0° to 180°, so each histogram bin corresponds to a 20° orientation interval. The computed cell histograms must be combined into a descriptor vector of the image. First, these cells should be locally contrast-normalized, due to the variability of illumination and shadowing in the image. This requires grouping the cells together into larger, spatially-connected blocks. Once the normalization is performed, all the histograms are concatenated in a single feature vector, the HOG descriptor. We use $[3 \times 3]$ cell blocks of $[6 \times 6]$ pixel cells with 9 histogram channels. The feature vector of each image, $I(H_j^i)$, is computed as its HOG descriptor, having 81 coefficients. It is expressed as:

$$V(H_j^i) = HOG(I(H_j^i)), \forall i \in [1, n], j \in [1, K]. \tag{11}$$

The perfect match for H_j^i is determined as the human object of the next frame corresponding to the minimum distance between feature vectors. Therefore, the entire video tracking process is modeled as following:

$$ind_i(j) = \arg \min_{t \in [1,K]} d\left(V(H_j^i), V(H_t^{i+1}) \right), \forall i \leqslant n, j \leqslant K \tag{12}$$

where d represents the Euclidean distance. So, each moving person tracked this way can be modeled as the following image object sequence:

$$Tr(j) = \left\{ H_j^1, H_{ind_1(j)}^2, \ldots, H_{ind_i(j)}^{i+1}, \ldots, H_{ind_{n-1}(j)}^n \right\}\Big|_{j \in [1,K]}. \tag{13}$$

The corresponding instances of the pedestrian in the initial color[1] video frames are then marked accordingly. A walking person tracking example is described in Fig. 2. The proposed object matching algorithm is applied on the same video sequence of the 3 consecutive frames analyzed for pedestrian detection in Fig. 1. Identified human objects, two for each video frame, are bounded by colored rectangles. The matching process is represented by arrows linking human objects from different frames and marked by the computed distances between their HOG-based feature vectors. The colored arrows, corresponding to lower distance values, indicate correct matches, while the white ones, marked by higher values, represent wrong person matches. The states of the same pedestrian are marked with the same color and linked by arrows of that color.

4. Experiments and method comparison

We have tested this human detection and tracking system on various video datasets containing moving people. Our numerous experiments, performed on hundreds video streams, provided satisfactory results. The proposed methods produce quite high person identification and tracking rates. Thus, the described person detection technique achieves an over 80% human object identification rate. The object detection rate, obtained by the proposed temporal differencing based algorithm, has a higher rate, of approximately 90%, but this rate is lowered by the performance of the skin segmentation technique. The described skin recognition algorithm produces robust skin-colored region identification, but can label some skin-colored regions, such as cloth pieces, as skin.

We have got some good values for the performance parameters: *Precision*, *Recall* and F_1. Our pedestrian detection technique produces few missed hits (undetected moving humans) and also few non-human objects wrongly labeled as persons. The proposed video tracking approach is also characterized by a high person matching rate, of approximately 90%. The resulted performance parameters are *Precision* = 0.90, *Recall* = 0.85 and $F_1 = 0.874$, meaning that very few false positives and false negatives are obtained by the tracking technique. We have described some of our many detection and tracking experiments in the two figures, Figs. 1 and 2. The following parameter values are used in the detection stage of our tests: threshold value $T = 30$ in (1) and $k < 3$ for the structuring element of the morphological process. The numerical experiments have been

[1] For interpretation of color in Figs. 1 and 2, the reader is referred to the web version of this article.

performed using MATLAB. We have implemented in MATLAB the algorithms for frame-difference based detection, skin locating and HOG-based feature extraction.

Method comparisons have also been performed. Thus, we have compared the performances of the pedestrian detection and tracking techniques presented here with the performances of some other well-known methods. As we have mentioned before, our approach uses the Histograms of Oriented Gradients in the tracking stage instead of using them in the detection phase. We have found that our technique produces better detection results than other frame difference-based and background subtracting approaches. Also, it provides comparable good results with the HOG-based human identification methods for non-occluded side view pedestrian detection [7], while running faster than those algorithms because of its lower computational cost. Our human detection method has also a much lower time complexity and execution time than detection approaches based on point detectors (like that based on Scale-Invariant Feature Transforms – SIFT), mean-shift clustering or Active Contours.

Unfortunately, our people detection system performs somewhat worse than the techniques using HOGs in the identification stage for front view pedestrian detection, rear view pedestrian detection and partially occluded pedestrian detection [25]. This pedestrian detection approach is also quite sensitive to undesired camera movements, because of the used frame-difference based procedure. The presence of a slight camera motion may affect seriously the entire moving object identification process. These are the major limitations of our proposed detection method.

The multiple human tracking algorithm proposed in this paper works considerably better than many other video tracking solutions. We have compared our HOG-based tracking results with those produced by other template matching based tracking approaches. We have found from numerical tests that pixel differencing based matching techniques, like those using SAD, MAD, MSD, correlation operations [17,24], color histograms or histogram intersections, execute somewhat faster than our HOG-based approach that has a higher computational complexity, but they also achieve poorer tracking results. Other tracking methods, like those using template matching with 2D Gabor filter based features or Exhaustive Search (ES), are clearly outperformed by our technique that runs faster and has a much higher tracking rate. Also, this HOG-based tracking algorithm has a lower time complexity than point tracking techniques, like those using Kalman filters, while producing comparable satisfactory results.

5. Conclusions

A novel automatic multiple person detection and tracking system for static camera video sequences has been proposed in this paper. The article brings original contributions in both human detection and object tracking areas. The first contribution is the multi moving object detection approach, based on a novel temporal differencing algorithm and some morphology-based procedures. The identified objects satisfying some proper body-shape related conditions are then categorized as persons. An important idea brought by this paper is using the presence of skin-colored regions to detect humans. The skin segmentation technique introduced for this purpose represents another important contribution of this article. The developed human tracking method based on a HOG-based object feature extraction and an object matching process is also an important achievement.

The performed detection and tracking experiments prove the effectiveness of the proposed technique. Its automatic character represents also an important quality that allows our technique to be successfully used for large video databases. Thus, video indexing and retrieval become some important application areas of it. This approach can also be applied in other computer vision domains, such as robotics, video surveillance and urban traffic monitoring.

While the described person detection algorithm works successfully for identification of non-occluded side view pedestrians [7], it performs somewhat poorer on occluded people [25], front-view or rear-view moving people. Also, our method is not appropriate for non-pedestrian person detection and tracking. It fails to detect static humans from videos and performs quite unsatisfactory for persons moving only some body parts. Humans not moving in upright position are also difficult to track. The sensitivity to undesired camera motions represents another major limitation of this detection and tracking approach. We intend to improve the presented technique, as part of our future research in this field, making it to perform better in the mentioned cases.

Acknowledgements

This work was mainly supported by the project PN-II-ID-PCE-2011-3-0027-160/5.10.2011 financed by UEFSCDI Romania. It was supported also by the Institute of Computer Science of the Romanian Academy, Iaşi, Romania.

References

[1] Ogale NA. A survey of techniques for human detection from video. Dept of Computer Science, University of Maryland, College Park. <www.cs.umd.edu/Grad/scholarlypapers/papers/neetiPaper.pdf>.
[2] Wren C, Azarbayjeani A, Darell T, Pentland A. Pfinder: real-time tracking of the human body. IEEE Trans Patt Anal Mach Intell 1997;19:780–5.
[3] Zhou JP, Hoang J. Real time robust human detection and tracking system. In: Proceedings of IEEE computer vision and pattern recognition (CVPR), vol. III; 2005. p. 149.
[4] Zhan C, Duan X, Xu S, Song Z, Luo M. An improved moving object detection algorithm based on frame difference and edge detection. In: Proc of IEEE 4th int conf image and graphics, Chengdu, China; 2007. p. 519–23.

[5] Liu Y, Haizho A, Guangyou X. Moving object detection and tracking based on background subtraction. In: Proceeding of society of photo-optical instrument engineers (SPIE), vol. 4554; 2001. p. 62–6.
[6] Schwartz W, Kembhavi A, Harwood D, Davis L. Human detection using partial least squares analysis. In: Proceedings of IEEE 12th international conference on computer vision; 2009. p. 24–31.
[7] Papageorgiou C, Poggio T. A trainable pedestrian detection system. Int J Comput Vis (IJCV) 2000:15–33.
[8] Dalal N, Triggs N. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), vol. 1; 2005. p. 886–93.
[9] Ghidary S, Nakata Y, Takamori T, Hattori M. Human detection and localization at indoor environment by home robot. In: Proceedings of IEEE int conf on systems, man, and cybernetics, vol. 2. Nashville, TN, USA; 2000. p. 1360–5.
[10] Hu X, Tang Y, Zhang Z. Video object matching based on SIFT algorithm. In: Proceedings of international conference on neural networks and signal processing; 2008. p. 412–15.
[11] Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. Int J Comput Vis 1998:321–31.
[12] Beleznai C, Fruhstuck B, Bischof H. Human tracking by fast mean shift mode seeking. J MultiMedia 2006;1(1):1–8.
[13] Peterfreund N. Robust tracking of position and velocity with kalman snakes. IEEE Trans Patt Anal Mach Intell 1999;21(6).
[14] Nergui M, Yoshida Y, Imamoglu N, Gonzalez J, Sekine M, Yu W. Human motion tracking and recognition using HMM by a mobile robot. Int J Intell Unmanned Syst 2013;1(1):76–92.
[15] Wixon L. Detecting salient motion by accumulating directionally-consistent flow. IEEE Trans Patt Anal Mach Intell 2000;22(8).
[16] Xiaowei L, Qing-Jie K, Yuncai L. A feature fusion algorithm for human matching between non-overlapping cameras. In: Chinese conference on pattern recognition, CCPR '08; 2008. p. 1–6.
[17] Barbu T. An automatic face detection system for RGB images. Int J Comput Commun Control 2011;6(1):21–32.
[18] Barbu T. Gabor filter-based face recognition technique. Proc Roman Acad Ser A 2010;11(3):277–83.
[19] Costin M, Barbu T, Zbancioc M, Constantinescu G. Techniques for static visual object detection within a video scene. In: Polytechnic institute bulletin, automatics and computers, Tom LI (LV), Fasc. 1-4; 2005. p. 75–85.
[20] Barbu T. Novel automatic video cut detection technique using Gabor filtering. Comput Electr Eng 2009;35(5):712–21.
[21] Barbu T. Variational image denoising approach with diffusion porous media flow. Abstract and applied analysis. Hindawi Publishing Corporation; 2013. p. 8, Article ID 856876.
[22] Serra J, Soille P. Mathematical morphology and its applications to image processing. In: Proceedings of the 2nd international symposium on mathematical morphology (ISMM'94); 1994.
[23] Fleck MM, Forsyth DA, Bergler C. Finding naked people. In: Proceedings of european conference of computer vision, ECCV; 1996. p. 593–602.
[24] Lewis JP. Fast normalized cross-correlation. Vision interface; 1995.
[25] Wang X, Han TX, Yan S. An HOG-LBP human detector with partial occlusion handling. In: Proceedings of ICCV '09; 2009.

Tudor Barbu is Senior Researcher at Institute of Computer Science of the Romanian Academy, Iaşi, Romania, and coordinator of Image and Video Analysis research team of the institute. He has a PhD degree in Computer Science and has authored 2 books, 1 book chapter and over 75 articles published in recognized international journals and volumes of international scientific events.