Fast Dynamic Mosaicing and Person Following

Andrea Prati
DISMI - Univ. of Modena and Reggio Emilia
Via Amendola 4, 42100 Reggio Emilia, Italy

Fabrizio Seghedoni, Rita Cucchiara
DII - Univ. of Modena and Reggio Emilia
Via Vignolese 905/b, 41100 Modena, Italy

Abstract

A system for video surveillance purposes in wide areas based on active cameras, also capable to follow a person in the scene by keeping him framed, is presented. The proposed approach is based on the so-called direction histograms to compute the ego-motion and on frame differencing for detecting moving objects. It exploits post-processing and active contours to extract precise shape of moving objects to be fed to a probabilistic algorithm to track moving people in the scene. Person following, instead, is based on simple heuristic rules that move the camera as soon as the selected person is close to the border of the field of view. Experimental results on a live active camera demonstrate the feasibility of real-time person following.

1. Introduction

This paper proposes a method for moving people segmentation and tracking from a moving camera, particularly conceived for video surveillance. The method is designed to work in real time for creating a mosaic image of the whole scene (by registering overlapped images provided by successive frames of the active camera), detect and track moving people very quickly, and follow a selected person. *Person following* is intended as the task with which the system keeps the person framed by the current view by automatically moving the active camera. This task can be useful for many applications, from face recognition purposes to snapshot logging for post-analysis and information retrieval.

We propose a new method for fast ego-motion computation based on the so-called *direction histograms* [7, 3]. The method works with an uncalibrated camera that moves with an unknown path and it is based on the compensation of the camera motion (i.e., the *ego-motion*) to create the mosaic image and on the frame differencing to extract moving objects. Successive steps eliminate the noise and extract the complete shape of the moving objects in order to exploit an appearance-based probabilistic tracking algorithm. Person following, instead, is based on a quite intuitive method that moves the camera when the person is near to exit from the field of view of the camera.

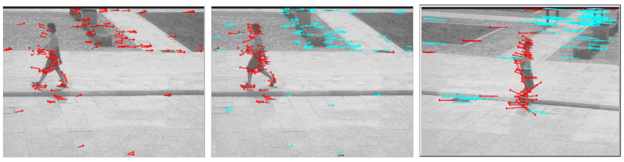
The segmentation of moving objects becomes more critical when the video is acquired by a moving camera with an unconstrained and a priori unknown motion. Proposals from single camera can be grouped into three classes: based on ego-motion computation, based on motion segmentation, and based on region merging with motion. The approaches in the first class aim at estimating the camera motion (or ego-motion) through the evaluation of the dominant motion with different techniques and models in order to obtain compensated videos and to apply algorithms developed for fixed camera (frame differencing, as in [2], or background suppression, as in [8]). In [5] Kang *et al.* define an adaptive background model that takes into account the camera motion approximated with affine transformation. Tracking of moving object is achieved by means of a joint probability data association filter (JPDAF). In methods based on motion segmentation the objects are mainly segmented by using the motion vectors computed at pixel level ([6]). The vectors are then clustered to segment objects with homogeneous motion. Finally, the approaches based on region merging with motion are hybrid approaches in which the objects are obtained with a segmentation based on visual features, and next merged on motion parameters computed on a region-level [4].

It is worth noting that most of the reported approaches are computationally very expensive and cannot meet real-time constraints (and those that meet them use either special-purpose devices or a set of limiting assumptions).

2. People Tracking from Moving Camera

Our approach for moving object segmentation from moving camera consists in two basic steps: first, the ego-motion is estimated and compensated to build a mosaic image and, second, frame differencing and post-processing are applied to extract the single moving objects.

The motion vectors of the current frame are extracted using a pyramidal implementation of the Lucas-Kanade algorithm (Fig. 1(a)). Then, they are clustered to find the dominant motion, that corresponds to the ego-motion assuming that the background is dominant over the moving objects.



(a) Motion vectors with LK (b) Clustered motion vectors (c) Clustered motion vectors with concordant motion

Figure 1. Extraction of the motion vectors

The clustering is performed with an innovative and fast process. It can be demonstrated that, for small pan and tilt angles and assuming the optical center fixed (good approximation for an active cameras), the camera motion model can be approximated with a pure translational model. With this hypothesis, a *direction histogram* containing all the directions of the extracted motion vectors is built (see Fig. 2). Let $\vec{\mathbf{mv}}(x,y) = (\rho(x,y), \alpha(x,y))$ be the motion vector computed at the coordinate (x,y). We define the direction histogram as $DH(\beta) = \#\{\vec{\mathbf{mv}}(x,y) | \alpha(x,y) = \beta\}$, with β ranging from 0 to 2π. A 1-D Gaussian filter centered on the histogram peak is applied on the histogram to eliminate motions different from the dominant one:

$$\widetilde{DH}(\beta) = DH(\beta) \cdot G(\mu, \sigma) \qquad (1)$$

where $\mu = \arg\max_{\beta} DH(\beta)$ and σ is a parameter set to 1 for most of the experiments.

The resulting histogram $\widetilde{DH}(\beta)$ allows to divide the motion vectors into two groups, one due to the camera motion (in cyan in Fig. 1(b)) and one due to moving objects (in red in Fig. 1(b)), and to compute the direction $\overline{\alpha}$ and amplitude $\overline{\rho}$ of the ego-motion by averaging the vectors retained by the Guassian filter:

$$\overline{\rho} = \arg\max_{\beta} \widetilde{DH}(\beta) \quad ; \quad \rho = \frac{\sum_{\rho \in R} \rho}{|R|} \qquad (2)$$

where $R = \left\{ \rho(x,y) | \widetilde{DH}\left(\alpha(x,y)\right) \neq 0 \right\}$.

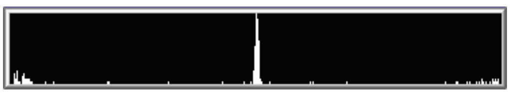
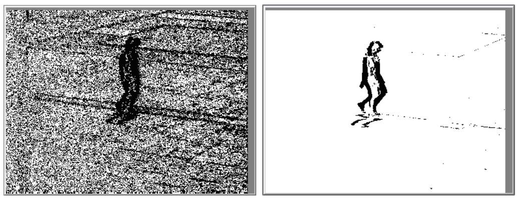


Figure 2. An example of direction histogram

This approach, though intuitive and simple, has proven to act very well, given that the above-mentioned hypothesis holds. For example, Fig. 1(c) reports the result in the case of a person moving with motion concordant with the camera: though some errors are present, accuracy is still acceptable.

Once the ego-motion is estimated, the current frame is registered (assuming a translational motion model) by compensating for the camera motion given by the vector $(\overline{\rho}, \overline{\alpha})$. The difference in the results performing frame differencing before and after the compensation is shown in Fig. 3. Moving pixels are indicated in black.



(a) Without compensation (b) With compensation

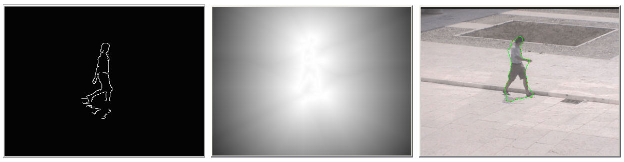
Figure 3. Frame differencing (a) without and (b) with ego-motion compensation

As evident in Fig. 3(b), the result provided by frame differencing are still far from being optimal, for both the noise due to imprecise image registration and the ghost of the moving objects. For this reason, post-processing steps must be used. Noise and small areas are removed by morphological operations, whereas ghosts are eliminated by merging information provided by a connected-components analysis and by a Canny edge detector: only edges with at least one point (in the 3x3 neighborhood) detected as moving are retained. Fig. 4(a) shows an example of retained edges.

Based on these information, the single moving objects are located. Their shape, however, is imprecisely extracted. Since the performance of the tracking algorithm heavily depends on the precision of the object's shape, a successive step is required. Since standard background suppression techniques are not suitable with our requirements (unknown path, uncalibrated camera, and real-time constraints), we employ a variant of the classical *active contours*, in which the energy is obtained with the following equation:

$$E_i = E_{cont,i} + \frac{E_{curv,i}}{2} + E_{dist,i} \qquad (3)$$

where $E_{cont,i}$ represents the contour continuity energy and $E_{curv,i}$ the contour curvature energy (the smoother the contour is, the lower the energy). As external energy, we modify the original proposal by considering the image obtained by applying the Distance transform to the image containing the edges retained by the post-processing. Examples of input edge image, external energy image and resulting snake are reported in Fig. 4. Finally, contour filling is employed to obtained a rough segmentation of the person's shape to be provided to the appearance-based probabilistic tracking proposed in [1], that is meant to be robust to occlusions.



(a) Edges (b) Energy (c) Snakes

Figure 4. Active contours.

Final mosaic image is constructed by superimposing the registered image on the mosaic and applying a simple alpha blending algorithm. Moreover, moving objects are not pasted onto the mosaic.

Once moving people are detected the system allows to select a single person to be followed, by moving the camera to keep him framed. In the current implementation of the system the "youngest" person (in the sense of that tracked by less time) in the scene is followed until he is visible. When he exits from the area the camera either follows the next "youngest" person, if any, or goes to a predefined position. Person following is achieved by moving the camera towards the person as soon as he approaches the limit of the current field of view.

3. Experimental Results

The system has been tested by using an active camera mounted in our campus. During the development of the system all the tests have been carried out offline, pre-storing videos from the camera and processing them offline. Now the system has been moved online, to work directly on the live camera, in order to provide to the camera the feedback necessary for person following.

Dynamic mosaicing has been evaluated by means of both qualitative and quantitative analysis. For qualitative analysis, a large set of videos has been taken with different illumination conditions, different number of people (from none to 5-6 simultaneously moving people), and different movements of the camera (only pan, only tilt, both pan and tilt). This analysis has demonstrated that, if the hypotheses hold, the system produces very good mosaic images, like those shown in Fig. 5. The only distortions appear at the top of the image where they do not affect moving object segmentation. Mosaic images have been also evaluated using a quantitative (i.e., objective) measure such as the PSNR. For example, the PSNR of the mosaic images reported in Fig. 5 with respect to ground truths (generated by exhaustively trying all the possible displacements and choosing that minimizing the error) is 40.82 dB.

Fig. 6 shows a sequence reporting some snapshots of the results for person following. The red bounding box identifies the person followed, while green ones identify other moving objects. The drawings on the bottom right corner of each image show the actual movement of the camera. It is worth noting that these results have been obtained with a completely unsupervised system working on live camera. It is evident that there are some imprecisions: for instance, on row 2, column 3, the second person is not segmented since it is very dark; on row 3, column 2, shadows are connected to the moving person; erroneous moving objects are detected on the column in the last row, columns 3 and 4. In particular, the last snapshot reports a wrong segmentation due to the presence, in the background, of much texture and to the closeup of the scene.

From the computational point of view, the system works in real time, with an average frame rate from live camera of about 8 fps, including also the person following task. Considering that the current acquisition device releases 12.5 frames per second, we can properly speak of "real time".

4. Conclusions

The extraction and tracking of moving objects from a moving camera are difficult tasks, especially under the constraints of unknown camera motion, uncalibrated camera, and fast system's response. This paper proposes a suitable solution that, given the typical hypotheses of an active surveillance camera, assures a good trade-off between speed and accuracy. Experimental results showed that, if the person does not move too fast with respect to the speed of the camera's moving head, real-time person following on live camera is feasible.

This work was supported by the project L.A.I.C.A. (Laboratorio di Ambient Intelligence per una Città Amica), funded by the Regione Emilia-Romagna, Italy.

References

[1] R. Cucchiara, C. Grana, G. Tardini, and R. Vezzani. Probabilistic people tracking for occlusion handling. In *Proceedings of Int'l Conference on Pattern Recognition*, volume 1, pages 132–135, Aug. 2004.

[2] R. Cutler and L. Davis. Robust real-time periodic motion detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):781–796, Aug. 2000.

[3] W. Freeman and M. Roth. Orientation histograms for hand gesture recognition. In *Proc. of Intl. Wkshp. on Automatic Face and Gesture Recognition*, June 1995.

[4] M. Gelgon and P. Bouthemy. A region-level motion-based graph representation and labeling for tracking a spatial image partition. *Pattern Recognition*, 33:725–740, 2000.

[5] J. Kang, I. Cohen, and G. Medioni. Continuous tracking within and across camera streams. In *Proceedings of IEEE Int'l Conference on Computer Vision and Pattern Recognition*, volume 1, pages I–267 – I–272, 2003.

Figure 5. An example of mosaic image.

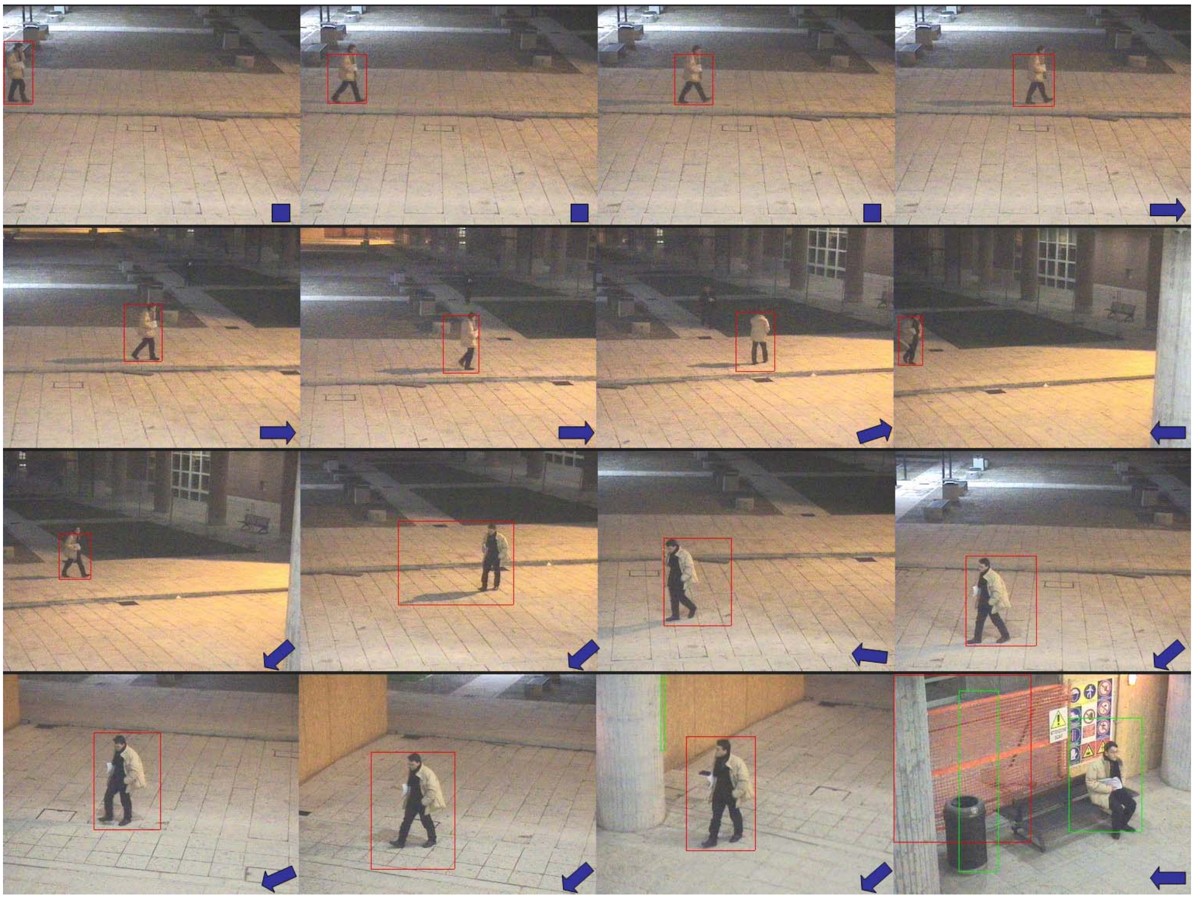


Figure 6. Snapshots from a live sequence with person following.

[6] K. Lee, S. Ryu, S. Lee, and K. Park. Motion based object tracking with mobile camera. *Electronics Letters*, 34(3):256–258, Mar. 1998.

[7] R. K. McConnell. Method of and apparatus for pattern recognition, u. s. patent no. 4,567,610, Jan. 1986.

[8] H. Sawhney and S. Ayer. Compact representations of videos through dominant and multiple motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):814–830, Aug. 1996.