REAL TIME FOREGROUND OBJECT DETECTION USING PTZ CAMERA

Lionel Robinault[1, 2], Stéphane Bres[3] and Serge Miguet[1]
[1] *LIRIS - Lyon2 University*
[2]*FOXSTREAM - LIRIS, Lyon2 University*
Bat. C, 5, Av. Pierre Mendès France
69676 Bron cedex – France
[3]*LIRIS, INSA Lyon*
Bat. Jules VERNE
69 621 Villeurbanne

Keywords: PTZ camera, Background/foreground detection, Gaussian mixture, Image registration.

Abstract : An important research is done to exploit the characteristics of PTZ cameras. These cameras allow motorized cover a wide field of view. A classic application of these cameras is to image mosaicing. But they can also be used to track moving objects. In this paper, we present an original approach for performing the registration, adapted to the case of central projection and a background subtraction algorithms for these cameras. The background image is iteratively updated and only on the part "seen" by the camera. We have experimented different segmentation algorithms using our background modeling technique and this approach makes it possible object tracking in real time for PTZ cameras.

1 INTRODUCTION

The goal of this paper is to present is to detect in real time the foreground objects from a moving camera PTZ. Most solutions described in the literature (Kang 03, Migdal 05, Bevilacqua 06) requires as a first step, create a complete panorama of the scene. This panorama is the modeling of background. During the operation, acquired images are projected onto the panorama. Moving objects are segmented from the difference between the panorama and the projection of the current image. This approach leads to many problems. Among other things, the acquisition time of the first stage and the size memory needed to store the panorama without loss of information. However, the most important is the time between the construction of background model and the acquisition of the current image. This problem is even more sensitive outdoor lighting that changes regularly.

In this paper we present a robust background modeling method adapted to PTZ cameras and does not require the creation of such a mosaic. The additional interest of our approach is the reduction of processing time, in order to deal with real-time constraints. The first step in our approach relates to the image registration. We propose a fast image registration method adapted to the specific case of central projection. The second step is to update a background image corresponding only to the field of view (FOV) of the camera at time t. The rest is erased from the memory.

This article is structured as follow: in the next section we present the state of the art and our approach of image registration. In the section, we propose an generalization of background modeling method adapted to PTZ cameras. Then in section 4 we present our experimental results. The conclusion and the perspectives are presented in section 5.

2 IMAGE REGISTRATION

2.1 State of the Art

Although many solutions have been proposed for building panoramas, achieving high quality mosaics

Robinault L., Miguet S. and Bres S. (2009).
REAL TIME FOREGROUND OBJECT DETECTION USING PTZ CAMERA.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 609-614
DOI: 10.5220/0001796806090614

in real time remains a very challenging task. The approaches can be classified according to the complexity of the model. Moreover, we can distinguish local *vs* global methods and direct *vs* feature-based approaches. Regarding model complexity, Bhat and al. (Bhat 00) use a simple translation motion models for motion segmentation with a PTZ camera. However, this assumption is only fulfilled for small tilt angles. More complex motion models are thus generally proposed, such as rigid, affine (Szeliski 97, Brown 03) or general projective models (Bevilacqua 05). In addition, most cameras deviate from a real pin-hole model due to radial distortion which becomes more prominent for shorter focal lengths, and some approaches (Sinha 04) propose to compensate it.

Local approaches aim at determining the model's parameters for each couple of successive frames, and consists in a frame to frame (or pairwise) registration. They are computationally efficient but this strategy introduces small alignment errors to accumulate. In particular, these errors become more evident when a video sequence returns to a previously captured location (problem known as "looping path"). Global approaches (Szeliski 97, Brown 03) formulate the registration problem in order to solve for all of the camera parameters jointly, i.e. by requiring that the ends of a panorama should join up. These kinds of exact optimization schemes are most of the time not compatible with real-time purpose, thus making global methods suitable mainly for batch computation.

Direct (or intensity-based) methods (Szeliski 97 , Sinha 04) attempt to iteratively estimate the camera parameters by minimizing an error function based on the intensity difference in the area of overlap. This can be achieved by computing the sum square difference (SSD) or ZSSD, the correlation coefficient (CC), the mutual information (MI) and the correlation ratio (RC). Szeliski and Shum (Szeliski 97) propose to estimate the registration homography by iteratively updating a correction matrix using the SSD. They use an affine model, but claim that their general strategy can be followed to obtain the motion parameters associated with any other motion models (perspective or even including radial distortion). In addition, they apply global alignment to the whole sequence of images, which results in an optimal image mosaic. Direct methods have the advantage that they use all of the available data and hence can provide very accurate registration, but they depend on the fragile "brightness constancy" assumption, and being iterative require initialization. Feature-based methods (Bevilacqua 05, Brown 03) start by establishing correspondences between points, lines or other geometrical entities for estimating the camera parameters. For example, Bevilacqua et. al (Bevilacqua 05) suggest to match current frame features (corners) to the background mosaic using the KLT tracker. They make use of a generic projective model, and propose to overcome the "looping path" problem with a feedback registration correction compatible with real-time requirements. In their approach no *a priori* information regarding the camera parameters or signals (pan/tilt angular movements). Thus, they use a histogram specification technique (Azzari 06) to manage automatic camera exposure adjustments (e.g. AGC) and environmental illumination changes (e.g. daytime changes). Brown and Lowe (Brown 03) propose to match SIFT features between all of the input images to form the panorama. They make use of an affine transformation model that they justify by the partially invariance of SIFT descriptors under affine change. They use a RANSAC algorithm as a probabilistic model for image match verification, in order to discard outliers for the parameters estimation. Finally, they use bundle adjustment (Triggs 00) as a global registration scheme to solve for all of the camera parameters jointly. Although the approach is efficient, and is able to automatically images being part of the mosaic, the panorama computation requires 83 seconds on a 2GHz PC.

2.2 Registration Problem Formulation

Mapping the current frame into a common reference coordinate system consists in determining the transformation between the acquired image I and the previously built panorama P, *i.e.* finding the homography between I and P. An homography is defined as a transformation between two projective planes. An exhaustive review of the projective transforms is beyond the scope of the paper, and the reader can refer to (Faugeras 93).

Projection Model. Using homogeneous coordinates, the homography corresponds to a linear transform that can be represented using a 3×3 matrix multiplication H. Denoting $X = (u,v,1)^T$ the coordinates of a point P_t in the current image I, the homography H maps P_t to $P'_t \in P$, whose coordinates are $X' = (u', v', w')^T$:

$$\begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} \approx H \times \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \approx \begin{pmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{pmatrix} \times \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \quad (1)$$

where ≈ indicates that equation 1holds up to a scale factor. The equation 1 gives the general form of a homography, with eight free parameters. However, the PTZ cameras constitute a special case. For example, we can assume that the camera's center of rotation is fixed and coincides with the center of projection while it is rotating and zooming. Such an assumption is valid, when the PTZ camera is used outdoors or in large environments where the shift of the camera center is small as compared to its distance to the observed scene. In that case, using a simplified model removing geometrical or chromatic distortions, the projection can be expressed as follows:

$$H = K_I \cdot R_{\varphi I} \cdot R_{\theta I} \cdot R_{\theta P}^{-1} \cdot R_{\varphi P}^{-1} \cdot K_P^{-1} \quad (2)$$

R_θ and R_φ being the rotation matrices in function of the pan and tilt angles, and K being the simplified matrix of the intrinsic parameters of the model:

$$K = \begin{pmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

where f_u and f_v correspond to the focal distance, given in pixel unit for the axes u and v, (u_0, v_0) is the projection center in the image plane.

Homography Estimation. Considering the two images I and P that have to be aligned, the registration problem can thus be formulated as estimating the homography $\widetilde{H}$ fulfilling the following equation:

$$\widetilde{H} = \text{argmin}_{H \in E} D(P, H(I)) \quad (4)$$

where E is the space search related to the homography parameters, and D is a dissimalirity measurement between P and $H(I)$. Solving the registration problem is thus two-fold. Firstly we have to define a image similarity measurement adapted to our context. Secondly we must specify how the minimization stated in equation 4 is carried out.

2.3 Our approach

Our problem consists in searching a homography between two images. In the state of the art, we have presented two classes of technique: intensity based and feature based methods. In the case of intensity based methods, the algorithm of minimization is fast however the evaluation of the cost function is slow. For the feature based methods, the research of interest point is fast but the computation of interest point features and matching of points is slow. We propose to mix the two approaches ie minimization algorithm and extraction of interest points.

As indicated in equation 4, we need to define a measure of dissimalirity. Usually, the cost function used is the sum square difference (SSD) measure or equivalent. The SSD measure is calculated between all pixels of image. For accelerate the computation time, we propose to use a cost function based on the position of the interest point. The first step consists to calculate the interest points (Harris 88) in two images. The interest points are calculated once at the beginning of the algorithm. At each iteration, we apply the transformation matrix at all points of I. The cost function is the sum distances between all points of P and the nearest point of I after transformation.

$$D = \sum_j \min\left(\left\| p_{i,P} - H'_2 \cdot p_{j,I} \right\| \right) \quad (5)$$

To optimize the computing time and avoid seeking the minimum distance between each point, we define a search area for each point of P. This search area is defined by the a priori knowledge (Fig1.)

There are many methods to find the minimum in a search space, such as simulated annealing and genetic algorithms. These methods are universally acknowledged to be less sensitive to local minima. However, the tests that we have done have shown that the number of intermediate solution is more important. For the mimization algorithm, we have choice a simplex method. The simplex method, introduced by Nelder and Mead in 1965 (Nelder 65), is now well-known optimization scheme applicable in a high-dimension space. It is based on the use of a polyhedron which dimensions are n+1, n being the unknown parameters to be determined. Each iteration updates the polyhedron in order to estimate the minimum of the cost function.

Moreover, compared to the simplex method, the conditions of stops on two other methods are more difficult to determine. The choice of the simplex method is therefore fully justified. In our application, the homography has five free parameters, as stated in equation 2. If none of these parameters is known, the simplex polyhedron shall have six vertices. If the parameters of the panorama

P are known or calculated at time t-1, only three parameters of image I have to be computed. The simplex is thus a tetrahedron.

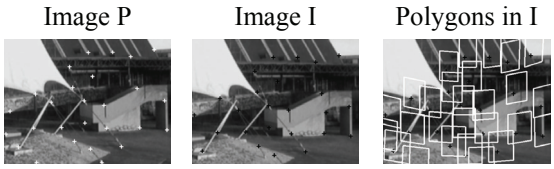

Figure 1: Search space for φp =45°, f_p=*830*, $\Delta\theta$=3°, $\Delta\varphi$ =3° and and Δf = *100*.

3 FOREGROUND SEGMENTATION

3.1 State of the Art

Several authors (Bhat 00, Kang 03) generate a preliminary complete (or partial) panorama of the scene. Then they projecting the current image in the panorama. There are several representations of panoramic images. One of them is to project all the images on a cylinder. This is the solution used in (Bhat 00).

However, making a complete panorama of the scene is particularly expensive in terms of memory. To store all of the scene without losing any information, it is necessary that the minimum size of each face of the cube is equal to twice the focal length expressed in pixels. For example, take a focal length corresponding to 800 pixels. For a color image, the required memory space is equivalent to 1600^2 x 3 x 6 or approximately 45 MB. If we use an algorithm based on Gaussian mixture, a minimalist solution requires 3 x 16-bit integers by Gaussian and it takes a minimum of two Gaussians. The memory is then 540 MB. The memory size is not the only limiting factor. For the background model to be meaningful, it is necessary to minimize the time for modeling the background as well as the computing time of the difference between current image and the background. If this time is too long, several factors make difficult to extract the moving objects. The change of brightness is also a factor. To continuously update the panorama is not a good option. The solution that we propose is, therefore, to model only the part of the background that is viewed by the camera in the current image.

Several approaches have been proposed for background modeling. The goal of this article is not to make a complete presentation of these methods, but we can cite three main families. The background image can be simply built from the previous frame or from a sliding average on previous images (Perner 01, Haritaoglu 00). The solution that seems to give the best results according to the bibliography is the method of Gaussian mixtures (Stauffer 99, Lee 05). We will enter with more details into these different methods.

3.2 Our Approach

The first step was to determine the transformation matrix between the current and previous images. We apply the transformation matrix to the background image I_f calculated at t-1 that we subtract from the current image I_c to obtain the map of foreground pixels I_m.

$$I_m = I_c - H \cdot I_f \quad (6)$$

There are several ways to calculate a background image. In this article we limit ourselves to one type of algorithm used by several authors (Stauffer 99, Lee 05). They model the change of each pixel in the image over time by using several Gaussian distributions represented by an average and a standard deviation. This method is commonly known as "Gaussian mixture". The number of distributions used for background modeling depends on the complexity of the background movements. The format of the article does not enable us to look further into the discussion on the relevance of this model and its parameters. For more information, the reader will be able to read the article of Stauffer (Stauffer 99). The tests which we carried out show that 3 distributions are generally necessary.

In the case of PTZ cameras, our approach consists in applying the transformation matrix to the different parameters (average, standard deviation) of the pixels of the background image. The distribution of each Gaussian can be accomplished by a bi-linear interpolation. Our approach makes it possible to use the transformation matrix on the background image and to put that back in the context of fixed cameras. We may use all classic algorithms of segmentation and identification of motion objects.

It is however important to notice that our approach does not allow us (under certain conditions) to segment all the moving objects. Indeed, the size of the background image being the same as that of the current image, we lose some information. That is, the area of the background image that was present on the previous image and who has disappeared with the movement of the

camera. It does not really matter because the camera movement is mostly linear in time and will therefore continue in the same direction. The camera does not change direction any time. What is more problematic is that a part of the background image is not available. Is the area of the current image that was not present in the previous image .

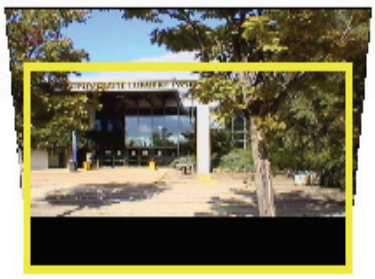

Figure 2: Background image projection on the current image.

In our example (Fig.2), we applied on the background image the transformation matrix that corresponds to the shooting parameters of the current image. The background image is projected in the plane of the current image. The rectangle shows the position of the current image in the plane. We can notice that a small part of the background image is outside of the rectangle. This is the part of the background that is lost. A black area appears inside the rectangle. This is the part that could not yet be analyzed by lack of modeling. In the above example we have voluntarily simulate a major movement in order to illustrate our point.

4 RESULTS

The following sequence (Fig.3.) corresponds to a real case. The camera could not give us a reliable position measurement, we used our registration technique. We present some images from the movie with the binarization results. The first column is the image acquired by the camera. In this example, the camera is rotating pan in the trigonometric direction. The overall scene is moving. In this scene, a pedestrian is also moving. The second column is to magnify the person in motion. Other columns correspond to the binarization of various methods. Column (WR) is the result of the difference, after binarization, between the current image and projection of the previous image in the plan of the current image. The projection matrix is estimated with our registration method. The column (CP) is the result of our background model but by using the parameters of the camera to calculate the projection matrix. Column (OA) is our approach (ie. image registration + mixture of gaussian). Compared to WR, our approach shows the contribution of our background model. In the case of CP, if the camera parameters were precise, the results would be comparable with our approach. However, will traditional PTZ cameras are not precise. For example, on the camera Sony RZ25P, the information of position is updated once time by second. If the positions taken by the camera are not just, the object segmentation is not perfect. Our approach helps to properly segment the pedestrian.

These tests were carried out on a laptop - HP Pavilion equipped with a 1.8GHz AMD processor and 1GB RAM. The computing times for 704 x 576 pixels images are 22ms for Gaussian mixtures. They make possible object tracking in real time.

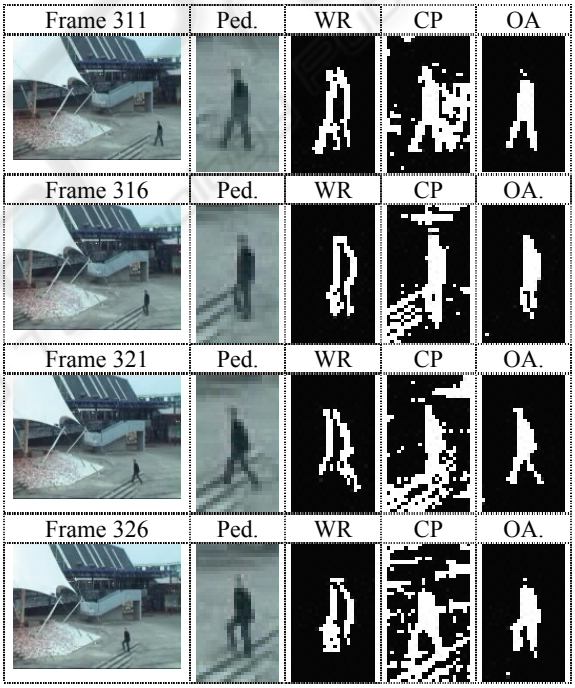

Figure 3: Real case sequence.

5 CONCLUSIONS

In this article we have presented a method for real-time background substraction adapted to PTZ cameras. The method we propose is not intended to achieve a robust panorama. It helps, however, to quickly calculate the projection between two successive frames of a video camera PTZ moving.

After the projection of the image J in the background of image I, the difference between the two images permitted the computation of the motion map. The best results are obtained with mixtures Gaussian. With the image registration, the computation time of the motion map is 29ms. The computation times reduced our method allows computing time available for other treatments, such as segmentation. Another advantage of our method is that it is less sensitive to changing light conditions. The brightness changes are immediately integrated as in the case of a fixed camera.

REFERENCES

Azzari, P., Bevilacqua, A., 2006, Joint spatial and tonal mosaic alignment for motion detection with ptz camera, *ICIAR06*. Vol. II, 764-775

Bevilacqua, A. and al, 2005, An effective real-time mosaicing algorithm apt to detect motion through background subtraction using a ptz camera. 511-516

Bevilacqua A., Azzari P, 2006, High-Quality Real Time Motion Detection Using PTZ Cameras, *AVSS '06*, p. 23

Bhat, K.S., Saptharishi, M., Khosla, P.K., 2000, Motion detection and segmentation using image mosaics, *IEEE International Conference on Multimedia and Expo*, 1577-1580

Brown M., Lowe D.G., 2003, Recognising panoramas, *ICCV 03*, 1218

Faugeras O., 1993, Three-Dimensional Computer Vision (Artificial Intelligence). *The MIT Press*

Haritaoglu I., Harwood D., Davis L.S., 2000, real-time surveillance of people and their activities, *IEEE Transaction On PAMI* , Vol. 22, pages 809–830.

Harris C., Stephens M., 1988, A Combined Corner and Edge Detector, *Proceedings of The Fourth Alvey Vision Conference*, 147-151

Kang S, Paik J., Koschan A., Abidi B, 2003, Real-time video tracking using PTZ cameras, *Proc. of SPIE 6th International Conference on Quality Control by Artificial Vision*, Vol. 5132, pp. 103-111

Lee D.S. 2005, Effective Gaussain Mixture Learning for Video Background Substraction, *IEEE Transaction On PAMI* vol.27, no. 5.

Migdal J., Izo T., Stauffer C., 2005, Moving Object Segmentation Using Super-Resolution Background Models, *OMNIVIS*

Nelder J.A., Mead R., 1965, A simplex method for function minimization. *The Computer Journal (7)*, 308-313

Perner P., 2001, Motion Tracking of Animal for Behavior Analysis, *Visual Form 2001*, Springer Verlag 2001, pages 779-787.

Sinha S., Pollefeys M., Kim S., 2004, High-resolution multiscale panoramic mosaics from pan-tilt-zoom cameras, *Indian Conference on Computer Vision, Graphics and Image Processing*. 28-33

Szeliski, R., Shum, H., 1997, Creating full view panoramic image mosaics and environment maps, *SIGGRAPH 97*

Stauffer C., Grimson W., 1999, Adaptative background mixture models for real-time tracking, *CVPR99*.

Triggs B., McLauchlan P., Hartley R., Fitzgibbon A., 2000, Bundle adjustment – A modern synthesis, *Vision Algorithms: Theory and Practice*. LNCS. Springer Verlag, 298-375