A Visual Object Counter Based on a Multimodal Data Association for Embedded Systems with Limited Computational Power

Carlos R. del-Blanco, Fernando Jaureguizar and Narciso García
Grupo de Tratamiento de Imágenes, ETSI Telecomunicación, Universidad Politécnica de Madrid, Spain

ABSTRACT

A low complex but highly-efficient object counter algorithm is presented that can be embedded in hardware with a low computational power. This is achieved by a novel soft-data association strategy that can handle multimodal distributions.

1. INTRODUCTION

Object counting is a crucial task in many practical applications such as urban planning, management of access points, traffic control, surveillance, etc. For this purpose, visual based systems offer many advantages over other technologies: the reduced cost of the equipment, and the low restrictions about the installation. This last advantage has been possible thanks to the latest tracking algorithms that are able to reliably track objects in complex situations with occlusions, which arise from camera points of view that are far away from the ideal azimuthal position. However, they have certain restrictions that can be inappropriate for some situations: the calibration of the camera [1, 2], and the high computational cost of the involved data association techniques [3, 4], needed to tackle the problems of multiple objects and occlusions.

In this paper, a novel Bayesian counter is presented that uses an uncalibrated camera to track multiple objects. This is accomplished by a soft data association technique that is able to handle multimodal data association distributions, unlike other soft-association techniques such as the Joint Probabilistic Data Association Filter (JPDAF). This feature allows to deal with object occlusions effectively. On the other hand, the complexity of the proposed algorithm is lower than that of hard-association strategies, making it suitable for low computational power devices.

2. BAYESIAN OBJECT COUNTER

The object counter is based on a Particle filter model that approximates the posterior probability over the flow of moving objects as a weighted sum of samples $\mathbf{x}_k^i$

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^{N_s} w_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i), \quad (1)$$

where N_s is the number of samples, $\mathbf{x}_k^i$ are the samples containing dynamic and positional information of all the objects in the time step k, and w_k^i are the samples weights.

This work has been supported by the Ministerio de Ciencia e Innovación from Spain under the project TEC2010-20412 (Enhanced 3DTV).

The samples are drawn from an importance sampling distribution $q(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k)$ that tries to approximate the actual posterior probability. In our model, the object dynamics is used as importance sampler

$$\mathbf{x}_k^i \sim q(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i, \mathbf{z}_k) = p(\mathbf{x}_k^i|\mathbf{x}_{k-1}^i), \quad (2)$$

The object dynamics is simulated by a piecewise constant velocity model, which is suboptimal, but very fast and efficient.

The sample weights are then computed to rectify the approximation of the posterior probability using the object detections. This is mathematically expressed as

$$w_k^i = w_{k-1}^i p(\mathbf{z}_k|\mathbf{x}_k^i), \quad (3)$$

where w_{k-1}^i are the sample weights in the previous time step, and $p(\mathbf{z}_k|\mathbf{x}_k^i)$ is the likelihood that uses the detections $\mathbf{z}_k$ at time k to weigh each sample $\mathbf{x}_k^i$. The detections are generated by the detector described in [5] that uses a background subtraction technique to estimate the moving objects in the scene. However, the output of a moving object detector can contain false detections and missing detections (not detected objects). The great challenge of the proposed model is to correctly estimate the real number of objects across the time taking into account the previous problems in the detection stage. Moreover, the correspondence between detections and moving objects is completely unknown, and this information is necessary to reliably estimate the current number of objects.

To work around the detection and association problems, a novel soft-association strategy is used, which is able to manage the multimodality of the data association process, unlike other soft-association techniques as JPDAF. This multimodality arises from imperfect detections and object occlusions [6]. This strategy is based on simulating the underlying detection distribution by a mixture of Gaussians whose means are the available detections. According to this, the likelihood conditioned to a specific object j is given by

$$p(\mathbf{z}_k|\mathbf{x}_{k,j}^i) = \sum_{r=1}^{N_d} \mathcal{N}(\mathbf{z}_{k,r}; \mathbf{F}\mathbf{x}_{k,j}^i, \mathbf{\Sigma}_d), \quad (4)$$

where N_d is the number of detections, $\mathbf{F}$ is a matrix that relates the positional information of the moving object and the detection, and $\mathbf{\Sigma}_d$ is a covariance matrix that expresses the variance of the detection process. Thus, the likelihood $p(\mathbf{z}_k|\mathbf{x}_k^i)$ can be expressed as a product of independent terms

$$p(\mathbf{z}_k|\mathbf{x}_k^i) = \prod_{j=1}^{N_o} p(\mathbf{z}_k|\mathbf{x}_{k,j}^i), \quad (5)$$

where N_o is the number of moving objects.

Using the previous equations, the posterior distribution over the flow of objects is expressed as

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^{N_s} w_{k-1}^i \prod_{j=1}^{N_o} \sum_{r=1}^{N_d} \mathcal{N}(\mathbf{z}_{k,r}; \mathbf{F}\mathbf{x}_{k,j}^i, \mathbf{\Sigma}_d). \quad (6)$$

The entrance of new objects is modeled by a Poisson distribution that expresses the probability that N_e new objects have entered in the scene in the last time interval. This approach assumes that a rough estimation of the average rate of new objects is known. The initialization of new objects is carried out by using those detections in $\mathbf{z}_k$ that have contributed least to estimate the position of the moving objects. The flow of objects $\mathbf{x}_k$ are then updated with the estimated new objects.

The exit of objects in the scene is modeled by a Gamma distribution that simulates the "time to death" of one object according to the last time that the corresponding likelihood term $p(\mathbf{z}_k|\mathbf{x}_{k,j}^i)$ was greater than a specific threshold. This reflects the fact that a detection is probably not generated by an object if the related likelihood value is too low. The flow of objects $\mathbf{x}_k$ are then updated by removing the objects that have exited.

Finally, the current object flow $\widetilde{\mathbf{x}}_k$ is obtained from the estimated posterior distribution $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ by means of the Maximum a Posteriori (MAP) estimator.

3. RESULTS

The proposed Bayesian object counter has been tested using several sequences of the PETS2006 dataset, which includes situations with a variable number of people in a train station from different points of view. Fig. 1 shows the estimated flow of people across the time, marked with a solid line, for a part of the sequence "S1-T1-C", camera number 3. Also, the real number of people is displayed with a dashed line. As it can be observed, the flow estimation is very similar to the ground truth, proving the high accuracy of the Bayesian counter. Also, notice that there is a little delay between the estimated and the ground truth flow, however this fact does not affect the total people count.

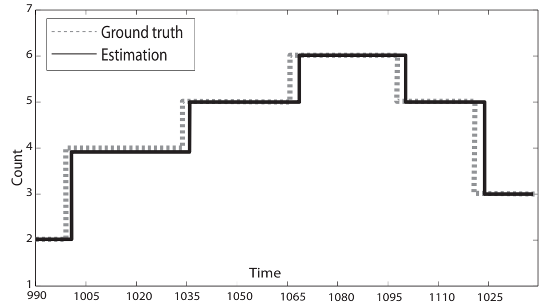

Fig. 1. Flow of people across the time.

General performance results are given in Table 1, where the accuracy of the people counter is given in percentage. From the obtained results, it can be concluded that the Bayesian counter has a high performance without the need of using a hard-association technique. This fact makes possible to embed the presented counting strategy in hardware devices with low computational power that need to fulfill real-time requirements.

Sequence name	Camera	Sequence length	Accuracy
S1-T1-C	3	3021 frames	93.5%
S1-T1-C	4	3021 frames	96.2%
S2-T3-C	3	2551 frames	94.1%
S2-T3-C	4	2551 frames	92.7%
S4-T5-A	3	3051 frames	94.5%
S4-T5-A	4	3051 frames	93.2%

Table 1. Accuracy of the Bayesian counter.

4. CONCLUSIONS

A Bayesian counter algorithm has been presented that can be embedded in hardware devices with a limited computational power, but with real-time requirements. This has been possible thanks to the design of a new soft-data association strategy that is able to manage multimodal distributions, unlike other similar approaches as the JPDAF. The performed results show a high accuracy in real situation with objects interactions.

5. REFERENCES

[1] K. Yam, W. Siu, N. Law, and C. Chan, "Effective bi-directional people flow counting for real time surveillance system," in *IEEE Int. Conf. Consumer Electronics*, 2011, pp. 863–864.

[2] O. Tuzel, F. Porikli, and P. Meer, "Learning on lie groups for invariant detection and tracking," in *IEEE Int. Conf. Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[3] Y. Lim and J. Choi, "Speaker selection and tracking in a cluttered environment with audio and visual information," *IEEE Trans. Consumer Electronics*, vol. 55, no. 3, pp. 1581–1589, 2009.

[4] C. R. del Blanco, F. Jaureguizar, and N. García, "Visual tracking of multiple interacting objects through rao-blackwellized data association particle filtering," in *IEEE Int. Conf. Image Processing*, 2010, pp. 821–824.

[5] C. Cuevas and N. García, "Moving object detection for real-time high-quality lightweight applications on smart cameras," in *IEEE Int. Conf. Consumer Electronics*, 2011, pp. 479 –480.

[6] R. Mohedano and N. Garcia, "Robust multi-camera 3d tracking from mono-camera 2d tracking using bayesian association," *IEEE Trans. Consumer Electronics*, vol. 56, no. 1, pp. 1–8, 2010.