In 15th Portuguese Conference on Pattern Recognition (RecPad 2009)
Aveiro, Portugal, October 2009

# Assessing Control Modalities Designed for Pan-Tilt Surveillance Cameras

Diogo Vicente  Jacinto C. Nascimento  José Gaspar
Institute for Systems and Robotics, Instituto Superior Técnico / UTL
dpvicente@gmail.com, {jan,jag}@isr.ist.utl.pt

## Abstract

*Many modern video surveillance systems encompass pan-tilt cameras due to the flexibility they provide in selecting the fields-of-view, as compared to using just fixed cameras. Although patent the great potential of using the pan-tilt cameras, one has to design pan and tilt controllers whose effectiveness directly impacts on the surveillance performance. In this work we propose a metric tuned to assess the effectiveness of such control designs, and show the theoretical estimation of the metric for the case of a one-object random-search controlling modality.*

## 1. Introduction

Depending on the purpose, surveillance can be understood more qualitatively as *awareness to novel events*, or more precisely as *tracking the trajectories* of moving objects or people walking. The first case is a *configuration finding* problem [4], where typically one wants information every time the surveyed area changes its default pattern e.g. due to a new object in scene. In the second case the solution involves mainly an *identification or data association* problem, in order to successfully track different objects.

Metrics were already proposed for accessing the performance of the basilar components of the image-based surveillance systems namely the segmentation algorithms. These metrics evaluate correct or false detections and object-splits, object-merges or both [2]. Metrics were also proposed for the higher levels of *configuration finding* and *tracking* methodologies [4]. Although being quite mature the outcome of [2] and [4], it focus on fixed cameras and in particular does not consider the nowadays, constantly growing number of, video surveillance installations encompassing pan-tilt cameras.

Surveillance with pan-tilt cameras involves not only video processing but also controlling the pan and tilt angles. Distinct controlling modalities imply distinct surveillance performances. The scope of this work focus on extending the metrics to assess these performances.

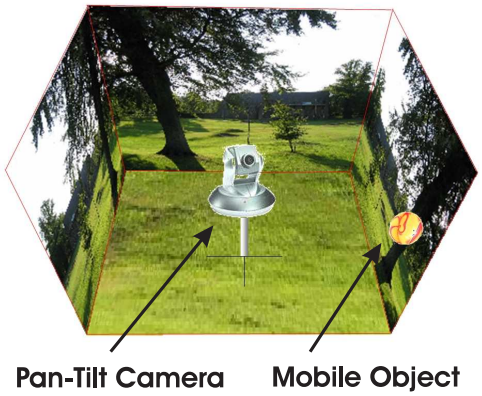


**Figure 1. Cube-based representation of the scene, not showing the top and front faces.**

## 2. Event detection and scene representation

There is a large variety of segmentation algorithms, i.e. algorithms doing intrusion / event detection in static scenarios. Some examples are *Basic Background Subtraction* (BBS), *Who? When? Where? What?* (W4) and *Single Gaussian Model* (SGM) [2]. The BBS, as the name indicates, simply compares a current image with a learned background. The W4 learns two backgrounds, the maximum and minimum expected gray scale intensities for each pixel, and detects differences whenever pixels have values outside of the learned ranges. In SGM each pixel is described by a mean and covariance which are updated recursively per frame and, based on this values, foreground and background pixels can be then identified. Notably, all these algorithms rely on the existence of a background representation of the scene.

There are also various manners to represent geometrically the background. For example one can use a planar mosaic, a cylinder, a sphere or a cube [1, 3]. In particular we select the cube based representation as it can handle a complete spherical field-of-view (FOV), $360^o \times 360^o$, which is not possible in the planar or cylindric mosaics, and

maps perspective images to/from the background using just homographies (as compared to using spherical mappings).

In order to assess surveillance methodologies, it is necessary to perform extensive testing and thus convenient to generate automatically ground truth information. It is therefore advantageous to build simulated setups. Figure 1 shows a cube based representation of a simulated scenario, having the pan-tilt camera in the center, which is able to survey objects moving within the 3D scene. In our experiments, this representation is maintained in two models, named the *operation model* which contains all the data, and the *background model* in which the mobile objects have been removed. The camera of the operation model captures images while surveying the test scenario, and the camera of the background model captures the corresponding images without the mobile objects, i.e. background images. This setup allows comparing test and background images as required by the segmentation algorithms.

## 3. Control modalities and performance metrics

In our work four modalities were considered for the control of the pan and tilt camera: *Random Search* (RaS), *Rotation Search* (RoS), *Local Search* (LoS) and *Local and Random Search* (LRS). Both RaS and RoS represent open loop algorithms where the camera would acquire images independently from the segmentation results. In RaS the sensing device acquires images while it is moved randomly (uniform distribution) within the pan and tilt limits. Hence, RaS requires very high operating speeds, to jump everywhere at anytime, and has expectedly a limited performance, as when it finds an object it does not try to keep it in the FOV. It is however an interesting control modality because of its simple statistical characterization. In RoS the camera is rotated (pan angle) with a constant step during image acquisition process, and thus searches systematically the scene, similar to RaS after a long time of operation.

LRS and LoS are closed loop algorithms, meaning that the segmentation results define the next orientation of the camera. Both algorithms try to center the object in the frame by computing and setting the pan-tilt angles when a detection is found. They differ when the object is lost, LRS starts commanding randomly the pan-tilt camera until an object is found, while LoS performs a local search around the last detection point before entering a random search mode.

An example of a very well known metric is the percentage of *Correct Detections* (%CD) in a sequence of $N$ images [2]:

$$\%CD = 100 \times \frac{\sum_{i=1}^{N} CD(I_i)}{\sum_{i=1}^{N} GT(I_i)} \qquad (1)$$

where $CD(I_i)$ denotes the number of correct detections (objects) found in the *i-th* image and $GT(I_i)$ is the ground truth number of objects in the image. In order to consider pan-tilt cameras, we propose instead using the percentage of *Events Found* (%EF):

$$\%EF = 100 \times \frac{\sum_{i=1}^{N} CD(I_i)}{\sum_{i=1}^{N} GT(I_i) + \sum_{i=1}^{N} GT(\bar{I}_i)} \qquad (2)$$

where $\bar{I}_i$ is an image based representation of the scene observable by the pan-tilt, but not accounted in $I_i$. Hence, the denominator of the fraction represents now all objects moving in the complete field-of-view of the pan-tilt camera, i.e. the number of non-background objects that can be observed by sweeping the full pan and tilt angle ranges.

In particular the RaS control modality allows obtaining a simple expression for the probability of finding one object by randomly sampling the pan-tilt complete field-of-view, and thus estimating the %EF metric. Assuming a punctual object and no image noise[1], then the %EF metric can be theoretically estimated by a ratio of solid angles:

$$\%EF_{RaS} \approx 100 \times \frac{\Omega_{cam}}{\Omega_{pan\times tilt}} \qquad (3)$$

where $\Omega_{cam} = 4 \arcsin(\sin(\alpha/2)\sin(\beta/2))$, is the solid angle of a perspective camera (pyramid) with $\alpha \times \beta$ (rad) FOV, and $\Omega_{pan\times tilt}$ is the solid angle corresponding to the complete FOV considering the full pan and tilt ranges. Therefore, $\Omega_{pan\times tilt}$ is the solid angle of a sphere minus the sphere caps not reached by the maximum tilting, $\Omega_{pan\times tilt} = 4\pi - 2\Omega_{cap}$. Each of the two non-reachable sphere caps can be represented by the solid angle of a cone with apex angle $2\theta$, i.e. $\Omega_{cap} = 2\pi(1 - \cos\theta)$ where $\theta$ is $pi/2$ minus the maximum tilt ($\tau_M$) and minus half the vertical FOV, $\theta = pi/2 - \tau_M - \beta/2$. In the case that the maximum pan angle ($\rho_M$) is less than $2\pi$, then $\Omega_{pan\times tilt} \leftarrow \rho_M/(2\pi) \times \Omega_{pan\times tilt}$.

## 4. Experiments

In order to illustrate the information introduced by the %EF metric, the control modalities have been applied in the simulated setup described in Sec.2, and assessed by both the %CD and %EF metrics. The mobile object (ball) moves around the camera, spanning a space larger than the FOV of the camera. Predictably, the open loop control modalities fail more detections of the object since they do not track it after finding it. This aspect is expected to be less severe as the FOV of the camera increases.

Figure 2 shows the performance metrics, %CD and %EF, for five FOV configurations. Each control modality has been tested on 500 images long sequences, using the

[1]Non punctual objects are considered by enlarging the camera FOV, and then the image noise can be mitigated by morphological processing.

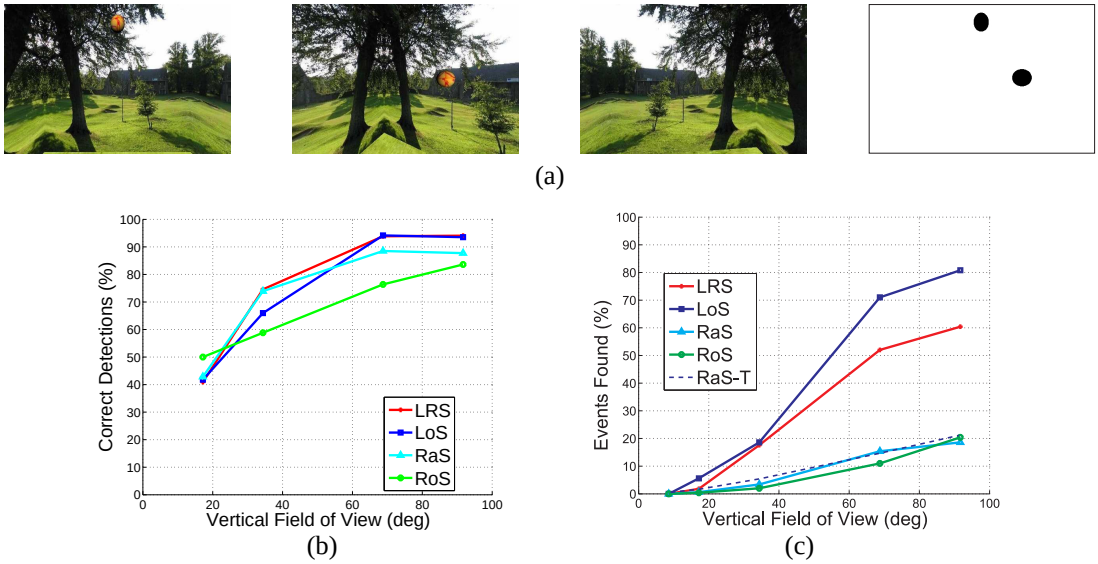


**Figure 2. Three sample images of the simulated scenario while the camera is panning to the left and the object is falling, and object detections superimposed on a single image (a). The %CD and %EF metrics versus the vertical FOV of the camera (b and c).**

BBS event detection methodology as in our synthetic scenario the W4 and SGM methodologies yield similar results. Both metrics confirm that all the control modalities tend to detect more times the mobile object when the FOV of the camera increases. Note however that the %CD metric does not show the expected clear distinction between open and closed loop control modalities. The interpretation is that the %CD metric does not count the objects that are out of the instantaneous FOV but, being in the vicinity of the camera, could be found (tracked) with a closed loop modality. The %EF metric effectively confirms the intuition that the closed loop control modalities are advantageous. Consistently, the theoretical prediction of the %EF for the RaS (labelled RaS-T in the plot) closely matches the experimentally observed results of the %EF and thus confirms the statistical significance of the realized number of experiments.

## 5. Final notes and future work

This article highlights the need of novel metrics for performance evaluation of surveillance systems encompassing pan and tilt cameras. While previously proposed metrics considered already false detections, object splits, merges, and both, and time/space evolutions such as identifying configuration vs tracking problems, most of the research was concentrated on fixed cameras. When considering pan-tilt cameras, one has also the objective of designing control algorithms that give a sense of the events happening in the complete scenario, in other words one desires to build surveillance systems that are more omni-aware. This work proposed a metric adjusted to evaluate such designs and showed its theoretical estimation for a case of random searching. Future work will focus on designing novel control modalities.

## 6 Acknowledgements

This work has been partially supported by the Portuguese FCT Programa Operacional Sociedade de Informação (POSI) in the frame of QCA III, the Portuguese FCT/ISR/IST plurianual funding through the POS Conhecimento Program that includes FEDER funds and the European Project FP6-EU-IST-045062 - URUS.

## References

[1] M. Brown and D. Lowe. Automatic panoramic image stitching using invariant features. *Int. Journal of Comp. Vision*, 74(1):59–73, 2007.

[2] J. C. Nascimento and J. S. Marques. Performance evaluation of object detection algorithms for video surveillance. *IEEE Transactions on Multimedia*, 8(4):761–774, 2006.

[3] S. N. Sinha and M. Pollefeys. Towards calibrating a pan-tilt-zoom camera network. In *Department of Computer Science, University of North Carolina at Chapel Hill*, pages 91–110, 2006.

[4] K. Smith, D. Gatica-Perez, and S. B. J. Odobez. Evaluating multi-object tracking. In *Comp. Vision and Patt. Recogn. - Workshops*, 2005.