An Eye-In-Hand Surveillance Camera For Human Patrolling Assistance Of Secured Areas

Fabio P. Bonsignorio

Heron Robots s.r.l.
Italy
Universidad Carlos III de Madrid
Spain

Abstract

In this paper a system for the assistance to human patrols of secured areas is described.
The system is designed as a patrolling aid driving the human supervisor attention where 'something is happening'.
A limit of current methods of surveillance of secured public or private areas by means of commercial fixed video cameras systems is that they require a continued human attention of all of the camera streamings. This require to engage a comparatively high number of people in the activity, and cannot guarantee a constant level of attention of all the people involved in it.
On a different respect research intelligent surveillance system based on the various methods proposed in the related literature do no not reach the assessment capabilities of properly trained human operators.
The system is conceived with a subsidiary approach leaving threat assessment and object recognition tasks to the human supervisors.
The system is seen as a robotic camera subsystem of a wider networked surveillance system involving systems and humans.
The camera intelligence software subsystem is designed as a saliency based attention system driven by the mutual information, between subsequent observations of the cells by which the observation scenario is divided.
This system integrates a simple parallel mechanism to drive the camera rotations mounted as a payload of a serial chain kinematics arm.
In comparison to similar saliency based attention approaches proposed in the literature the dynamical constraints of information metrics evolution due to the system mechanical configuration are considered, thus reducing computation needs.
A survey of surveillance systems state of the art technology from industrial and research standpoint is given and the limitations and opportunities of the different methods by themselves and in comparison to what we propose here are discussed.
The system mechanical and computing architecture and their underlying principles are described and critically analyzed.

Introduction

The surveillance (in particular during the night) of building and their surroundings and other secured areas (like metro station, warehouses, port docking areas are) are very often performed by video cameras fixed in carefully chosen position of the area under supervision.. The video feeds are usually recorded on tape or more recently on hard disk in digital format. These feeds are usually transmitted via analogue cable or data network to a remote station where one or more security operators, usually working on shifts, look at the video streams in order to detect anomalous situations and to assess the potential danger and severity of security breaches. Each operator look after a variable number of screens ranging from one to about ten. This security process organization has two major flaws. Ensuring a 24/7 surveillance requires 5/6 people

for group of screens. In the most favorable conditions 1 people every two security cameras (if an operator supervise ten cameras at a time and has longer surveillance shifts). This leads to comparatively high personnel costs even if the operator wages are kept low (which can negatively impact the skill level and motivation of the people performing the task). On the other end, in particular at night and at the end of the shifts, the attention levels of the patrolling people tend to decrease. This has led to propose (mainly as research prototypes although there are a few real world applications) ambient intelligence systems capable of automatically identify and classify potential intrusions.
As an example see [7].
These systems are stemmed by the lack of robustness and accuracy typical of many current technology AI solutions.
This papers describes a system design for a surveillance system which is intended as a tool to augment the human supervisor capability to detect security issues, while leaving the situation assessment duties to the human patrollers, with a scalable level of autonomy and a constant adaptation to the reconstructed level of attention of the user.
This is done implementing a saliency-based attention system based on mutual information between subsequent images.
This system integrates a simple parallel mechanism to drive the camera rotations mounted as a payload of a serial chain kinematics arm.
In comparison to similar saliency based attention approaches proposed in the literature the dynamical constraints of information metrics evolution due to the system mechanical configuration are considered, thus reducing computation needs.
This is meant of a subsystem constituted by a network of such eye-in-hand cameras integrated with fixed cameras and other mounted on various kind of mobile platforms.
In the following sections we will first give an overview of saliency based systems developed in order to understand natural attention processes in the primates and the humans. Then we will give an overview on Shannon information concepts. Eventually we will describe and discuss the proposed robotic system and we will give an idea about future directions of work.

Saliency-based models

Saliency-based attention has been proposed as an explanation mechanism to cut the computation load of scene analysis in primates and humans,[25]. Instead of processing with the same high level of detail all the scene only a few 'salient' areas of the images are analyzed. As a trade off the search of relevant visual features in the video stream is biased by the heuristics or other criteria adopted to identify the more relevant section of the video stream to analyze, see [11].
Primates' brains have evolved to effectively and efficiently identify salient sub-regions of the visual scene in real time in a way well adapted to human tasks. The bottom up high-lightening of sub regions of the images sufficiently different from their surroundings is usually biased by the tasks to be performed, [12, 17].
Although the general process of salience based scene analysis is widely accepted among primate and human neurophysiologists, there is a wide variation about the underlying mechanisms responsible for identifying 'salient' areas or features. A promising approach might be the so called AIM, Attention based on Information Maximization. These approach instead of hypothesizing an ad-hoc heuristics identify a general simple underlying mechanism: maximizing the Shannon information of the scene reconstruction. In [8] it is shown how the model is in good agreement with a broad range of observations.
In that paper visual saliency is equated to the amount of information on the scene subarea stored into a neuron or neuron network. This approach shows to be quite effective in predicting human fixation patterns and seems plausible from a biological perspective, in particular explaining some apparently counter intuitive behaviors observed in primates.
While other attempts to characterize the information content of a spatial location in the visual field have relied on the entropy of features as measured locally. Bruce and Tsotsos model the information content of a neuron as $y = -\log(p(x))$ where x is the firing rate of the neuron in question and p(x) the observation likelihood associated with the firing rate x. The likelihood of a given firing pattern of a neuron is predicted by the response of neurons in its support region. The approach is depicted in fig. 1.

Although a likelihood estimate based on a local window of image pixels appears to be a computationally hard problem as it requires the evaluation of probability density in high-dimensional spaces (e.g. 75 dimensions for a 5x5 RGB area), it is actually possible because the estimate is highly structured. The visual system actually exploits this property operating a compression.

A similar approach was proposed by Itti and Baldi [19]. In this case saliency is defined as related to 'surprise' modeled in term of an information metric based on Kullback–Leibler divergence between prior and posterior representations of visual content.

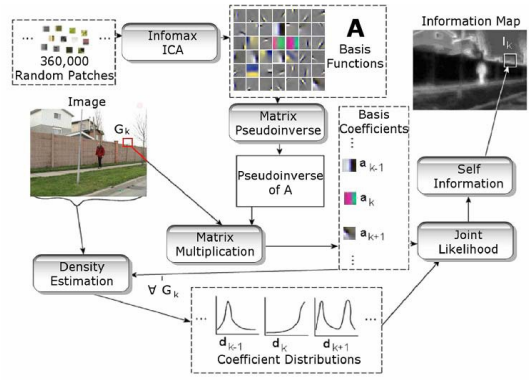

Fig. 1 Bruce and Tsotsos's saliency based model of attention, from [8]

Shannon Entropy and Related Metrics

In [1] Shannon proposed a measure of information in a distribution, which he called the `entropy'.

The entropy H(P) of a distribution P measures `the inherent uncertainty in P', or (in fact equivalently), `how much information is gained when an outcome of P is observed'.

More precisely, let us imagine an observer who knows that X is distributed according to P. The observer then observes X = x. The entropy of P stands for the `uncertainty of the observer about the outcome x before he observes it'. If you think of the observer as a `receiver' who receives the message conveying the value of X.

From this point of view, the entropy stands for the average amount of information that the observer has gained after receiving a realized outcome x of the random variable X.

Let X be a finite or countable set, let X be a random variable taking values in X with distribution P(X = x) = px. Then the (Shannon) entropy of random variable X is given by:

$$H(P) \doteq \sum_{x=X} p_x \log \frac{1}{p_x} \qquad (1)$$

Entropy is defined here as a functional mapping random variables to real numbers. In many texts, entropy is, essentially equivalently, defined as a map from distributions of random variables to the real numbers.

The entropy function can be derived in different ways. The two most common ones are the axiomatic approach and the coding interpretation.

It is interesting to know that the definition (1) can be derived from a small number of axioms:

1. $H(p_1,\ldots, p_N)$ is continuous in $p1,\ldots ,pN$.

2. If all the p_i are equal, $p_i = 1/N$, then H should be a monotonic increasing function of N. With equally likely events there is more choice, or uncertainty, when there are more possible events.

3. If a choice is broken in two successive choices the original H should be the weighted sum of the individual values H

It can be shown that a functional satisfying the above axioms, must have the form:

$$H(P) = k\sum_{i=1}^{n} p \log \frac{1}{p_i} \qquad (2)$$

With a similar approach it is possible to define the 'mutual information' between two random variable X and Y, which quantify the reduction uncertainty on X following a given outcome y of Y, as in (3)

$$I(X;Y) = -\sum_i \sum_j P_{XY}(i,j) \log \frac{P_X(i) P_Y(j)}{P_{XY}(i,j)} \qquad (3)$$

Detection System

Our system integrate a simple parallel mechanism to drive the camera rotations mounted as a payload of a serial chain.
The system is conceived with a subsidiary approach leaving threat assessment and object recognition tasks to the human supervisors.
The system is seen as a robotic camera subsystem of a wider networked surveillance system involving systems and humans.
The camera intelligence software subsystem is designed as a saliency based attention system driven by the mutual information, between subsequent observations of the cells by which the observation scenario is divided.
In our case we compute directly the mutual information between two subsequent observation identifying saliency with the local minima of mutual information.
In comparison to similar saliency based attention approaches proposed in the literature the dynamical constraints of information metrics evolution due to the system mechanical configuration are considered, thus reducing computation needs.
In our case, as it is supposed to occur in natural attention system, the computational load is strongly reduced by the fact that the scenes are structured and the mobility of the 'eye' is limited by the kinematic structure of the arm.

HW/SW Architecture

The system is made by an eye-in-hand camera mounted at the end of a 5 dof serial manipulation arm on a Stewart platform. The Stewart platform, [13], was originally designed, in 1965, as a flight simulator physical platform, and it is still widely used for that purpose.
The whole structure has a total of 11 degrees of freedom-
The 6 dof of the end effector (x,y,z and pitch, yaw and roll), allow to orient and position the camera with high flexibility, although by mounting at the end of a serial chain manipulator reduces the inherent stiffness and positioning accuracy typical of Stewart platforms (which only hold with reference to the end-effector).
The SW architecture is depicted in the following fig. 2.

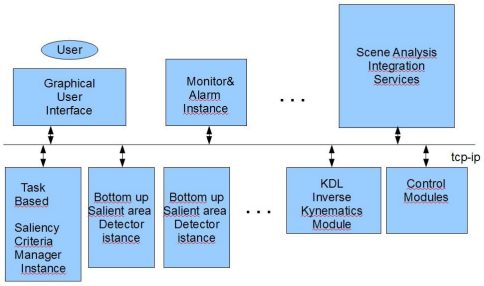

Fig. 2 SW Architecture

The system, developed in C++, is being developed in Webots, []. The executing platform is being implement on a parallel computing backbone.

Discussion and Future Work

The system shortly described here is bioinspired as it implements an attentional process model proposed to explain attention mechanisms in primates and humans. It differs from similar models proposed in literature as it directly compute the information metrics without relying on a artificial or natural neural networks. It, on the other hand, exploits computing parallelism and maximization of information measures as it is supposed to occur in natural attentional systems in the primates.
In order to draw conclusions it is necessary an extensive simulation (and 'field') experimental campaign and to define experimental protocols to properly compare with the described system and its natural counterparts.
It is thought that the exploitation of the Lie group subgroup representation in SE(3) of the system kinematics may help to reduce the computational loads and improve the control strategy. This require some theoretical work, which is ongoing-

References

1. Shannon,C.E.: The Mathematical Theory of Communication, Bell Sys. Tech. J. 27,379,623 (1948)
2. Kolmogorov, A.N.: Three approaches to the quantitative definition of information. Problems Inform. Transmission, 1(1):1-7, (1965)
3. Chaitin, G.J.: On the length of programs for computing finite binary sequences: statistical considerations. J.Assoc. Comput. Mach., 16:145-159,(1969)
4. Pfeifer, R.: Cheap designs: exploiting the dynamics of the system-environment interaction. Three case studies on navigation. In: Conference on Prerational Intelligence --- Phenomonology of Compexity Emerging in Systems of Agents Interacting Using Simple Rules. Center for Interdisciplinary Research, University of Bielefeld, 81-91, (1993)
5. Pfeifer, R. and Iida, F.: Embodied artificial intelligence: Trends and challenges. Embodied artificial intelligence, Iida et al. (Eds), LNCS/AI Vol. 3139, 1-26, Springer, (2004)
6. Allman, J., Miezin, F., McGuinness, E., :Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons. In visual neurons. Annual Review of Neuroscience 8:407-430, (1985)
7. Marcenaro, L., Marchesotti, L., Regazzoni, C.S.: Self-organizing shape description for tracking and classifying multiple interacting objects, Elsevier Image and Vision Computing, 24, 1179-1191, (2006)
8. Bruce, N., Tsotsos, J. K.: Saliency Based on Information Maximization. In: Advances in Neural Information Processing Systems, 18:155-162, (2006).
9. Cannon, M. W., Fullenkamp, S. C.: Spatial interactions in apparent contrast: inhibitory effects among grating patterns of different spatial frequencies, spatial positions and orientations. Vision Research 31:1985-1998, (1991)
10. Carmi, R., Itti, L.: Visual Causes versus Correlates of Attentional Selection in Dynamic Scenes. Vision Research 46(26):4333-4345, , (2006)
11. Crick, F.: Function of the thalamic reticular complex: the searchlight hypothesis. Proceedings of the National Academies of Sciences USA 81(14):4586-90, (1984)
12. Desimone, R., Duncan, J.: Neural mechanisms of selective visual attention, Annual Review of Neuroscience 18:193-222, (1995)
13. Stewart, D.: A Platform with Six Degrees of Freedom, UK Institution of Mechanical Engineers Proceedings 1965-66, Vol 180, Pt 1, No 15, (1965)
14. Frintrop,S., Jensfelt,P., Christensen, H. : Attentional Landmark Selection for Visual SLAM. In: Proc. IEEE International Conference on Intelligent Robots and Systems (IROS'06), (2006)
15.Hamker,F.H.:The role of feedback connections in task-driven visual search. In: Heinke,D., Humphreys,G.W., Olson, A. (Eds.), Connectionist Models in Cognitive Neuroscience. Springer Verlag. London, pp. 252-261, (1999)
16. Itti, L., Koch, C.,Niebur, E.: A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20(11):1254-1259, (1998)
17. Itti, L., Koch,C.: Computational Modeling of Visual Attention. Nature Reviews Neuroscience 2(3):194-203, (2001)
18. Itti, L.: Automatic Foveation for Video Compression Using a Neurobiological Model of Visual Attention, IEEE Transactions on Image Processing 13(10):1304-1318, (2004)
19. Itti, L., Baldi, P.: Bayesian Surprise Attracts Human Attention. In: Advances in Neural Information Processing Systems, Vol. 19 (NIPS*2005), Cambridge, MA, MIT Press, (2006)
20. Koch, C., Ullman, S.: Shifts in selective visual attention: towards the underlying neural circuitry. Human Neurobiology 4:219-227, (1985)
21. Navalpakkam, V., Itti, L.: Modeling the influence of task on attention, Vision Research 45(2):205-231, (2005)
22. Navalpakkam, V., Itti, L.: Search goal tunes visual features optimally, Neuron 53(4):605-617, (2007)
23. Saenz, M. , Buracas, G. T., Boynton, G. M. : Global effects of feature-based attention in human visual cortex. Nature Neuroscience 5(7):631-632,(2002)
24. Siagian, C., Itti, L.: Biologically-Inspired Robotics Vision Monte-Carlo Localization in the Outdoor Environment, In: Proc. IEEE International Conference on Intelligent Robots and Systems (IROS'07), (2007)
25. Tsotsos, J.K.: Is Complexity Theory appropriate for analysing biological systems? Behavioral and Brain Sciences 14(4):770-773, , (1991).