Affordances in Video Surveillance

Agheleh Yaghoobi*, Hamed R.-Tavakoli†, Juha Röning‡

*Electronics Laboratory, †Center for Machine Vision Research
‡ Computer Science and Engineering Department
University of Oulu, Finland
{agheleh.yaghoobi, hamed.rezazadegan, juha.roning}@ee.oulu.fi

Abstract. This paper articulates the concept of affordances use as the building block of an automated video surveillance system which learns and evolves over time. It grounds its arguments on the basis of a visual attention hardware and affordances.

Keywords: Surveillance, attention modeling, affordances

1 The problem scope

Video surveillance is an old demand influencing computer vision. Despite the recent impressive progress, there is still a long way to achieve a fully automated system and many of the prerequisites in this area require careful attention and are somehow a challenge, e.g., background subtraction [2], anomaly detection [3], and etc.

Traditionally successful commercial systems (e.g. SISTORE CX series from Siemens [9]) perform a centralized scene analysis in which violation of a series of predefined rules, which are usually imposed by an operator, trigger an alert. While it seems to be a long way to achieve having automated surveillance system which evolves and learns over time, the affordances theory [4] and visual attention modeling [12, 1] somehow promise to pave the way towards such an ultimate aspiration.

In this context, an automated framework is constrained by limited computational resources, volatile conditions of the environment (e.g. amount of crowd), the running site (e.g. a university campus or a factory), understanding the relation between the entities (i.e. scene understanding), and etc. Notwithstanding the difficulties, probably, one can still achieve a degree of automation by utilizing new concepts adopted from cognitive studies. Thus, the research question is: Can one utilize affordances to advance the surveillance to the next level?

2 Is affordances a solution?

The answer is not a straightforward affirmative phrase, neither a negative response. Although it is not the sole solution, it can be an important part of the answer by facilitating efficient scene processing. It possibly provides the necessities of an ontological-based surveillance platform which evolves over time. The following elaborates how one can implement such a system.

2.1 A rough sketch of the design

A practical approach for implementing such a platform consists of a combination of hardware and software units. Contrarily to the commercially available systems, the camera unit shall be updated to carry out part of the processing. Thus, an array of way more intelligent cameras will feed images and extra information, called meta-tags, to a central system. Meta-tags provide complementary information about the elements of the scene, e.g., it is a moving element, how contrasting the item is compared to its surrounding, and etc. These information are extracted using bottom-up visual attention models or salience modeling techniques, e.g. [6, 8], which are embedded in hardware. Afterwards. the video frame and the meta-tags are efficiently encoded [7] to be transferred to a central processing unit.

In the central unit, each video frame is processed using a series of contextual priors [10] and atomic or compositional rules imposed by predefined affordances. Atomic rules are defined as properties of an element independent of its behavior, e.g., *move-ability* defines if an element is able to move or not. On the other hand, a compositional rule consists of several atomic affordances at the same time, e.g., *aggressive movement* can be identified by existence of fast movement towards the site which requires identification of a moving element with particular movement pattern and characteristic.

Contextual priors are also important. In essence, they define the operation environment of the system, more accurately each camera unit. In such a system, two kind of priors exists, 1) excitery priors and 2) inhibitory priors. The first defines the existence of an element and its properties such as probable location. For instance, if a camera shall expect human presence in its field of view or vehicles and where should look for them. Contrarily, the inhibitory priors ban occurrence or existence of an element. In a nutshell, contextual priors ease the building process of an ontological tree [11] that helps understanding the environment.

The ontology evolves either via user interaction or recognition modules which identify the presence of elements and their interactions with the advent of affordances. While a user can alter both domain and conceptualization of the system, the recognition modules only affect changes in the domain (i.e. field of view of each camera). The automatic ontology enrichment and evolving is possible via graphical models in which a mapping between the ontology and an appropriate Bayesian network is derived, e.g. [5].

In the end, the outcome will be a set of hardware and software that vigilantly performs video surveillance, easily adapts to environment, and enhances over time. Also, a series of new affordances defined in the context of the entities of assigned task are expected. Eventually, a system, which integrates bottom-up visual attention techniques, contextual priors and affordances, will be implemented to derive ontological scene understanding in a less general scenario. In summary, the broad vision is a step toward convergence of cognitive sciences, electrical engineering and computer vision.

References

1. Borji, A., R.-Tavakoli, H., Sihite, D.N., Itti, L.: Analysis of scores, datasets, and models in visual saliency prediction. In: ICCV (2013)
2. Bouwmans, T.: Recent advanced statistical background modeling for foreground detection: A systematic survey. Recent Patents on Computer Science 4,(3), 147–176 (2011)
3. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection : A survey. ACM Computing Surveys 41(3) (2009)
4. Gibson, J.J.: The theory of affordances. In: Perceiving, Acting, and Knowing (1977)
5. Ishak, M.B., Leray, P., Amor, N.B.: A two-way approach for probabilistic graphical models structure learning and ontology enrichment. In: KEOD. pp. 189–194. SciTePress (2011)
6. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Anal. Mach. Intell 20(11), 1254 –1259 (1998)
7. Ma, T., Hempel, M., Peng, D., Sharif, H.: A survey of energy-efficient compression and communication techniques for multimedia in resource constrained systems. Commun. Surveys Tuts 15(3), 963–972 (2013)
8. Mancas, M., Riche, N., Leroy, J., Gosselin, B.: Abnormal motion selection in crowds using bottom-up saliency. In: ICIP. pp. 229–232 (2011)
9. Siemens: SISTORE CX, Configuration Manual. Siemens Building Technologies AG
10. Torralba, A.: Contextual priming for object detection. Int. J. Comput. Vision 53(2), 169–191 (2003)
11. Town, C.P.: Ontology based Visual Information Processing. Ph.D. thesis, University of Cambridge (2004)
12. Tsotsos, J.K.: A Computational Prespective on Visual Attention. The MIT Press (2011)