Research Article
Precision Security: Integrating Video Surveillance with Surrounding Environment Changes

Wenfeng Wang,1 Xi Chen,1 Guiwei Zhang,2 Jing Qian,3 Peng Wei,4 Boqian Wu,5 and Hongwei Zheng1

1CAS Research Center for Ecology and Environment of Central Asia, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Ürümqi 830011, China
2School of Management Engineering and Business, Hebei University of Engineering, Handan 056038, China
3Center for Geo-Spatial Information, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
4School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
5Key Laboratory of IOT Terminal Pivotal Technology, Harbin Institute of Technology, Shenzhen 518000, China

Correspondence should be addressed to Guiwei Zhang; zhangguiwei@hebeu.edu.cn

Received 22 July 2017; Revised 19 October 2017; Accepted 14 December 2017; Published 8 February 2018

Academic Editor: Roberto Natella

Copyright © 2018 Wenfeng Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Video surveillance plays a vital role in maintaining the social security although, until now, large uncertainty still exists in danger understanding and recognition, which can be partly attributed to intractable environment changes in the backgrounds. This article presents a brain-inspired computing of attention value of surrounding environment changes (EC) with a processes-based cognition model by introducing a ratio value $\lambda$ of EC-implications within considered periods. Theoretical models for computation of warning level of EC-implications to the universal video recognition efficiency (quantified as time cost of implication-ratio variations from $\lambda_k$ to $\lambda_{k+1}$, $k = 1, 2, \ldots$) are further established. Imbedding proposed models into the online algorithms is suggested as a future research priority towards precision security for critical applications and, furthermore, schemes for a practical implementation of such integration are also preliminarily discussed.

1. Introduction

Surveillance plays a vital role in maintaining social security and protecting infrastructure facilities of a country [1, 2]. But until now, there are still considerable uncertainties associated with danger understanding and recognition, especially for engineering-critical applications [3–5], which can be partly attributed to implications of environment conditions to video recognition efficiency of the surveillance system. It has been demonstrated that suitable model parameters in online algorithms and difficulty level of object detection tasks in different environments can be much different [6].

Surrounding environment changes as particular changes in backgrounds are also responsible for some significant but still unresolved issues in objects recognition and tracking [7]. Because the backgrounds cannot be well-characterized in uncontrolled environments changes, the surveillance video recognition becomes more intractable [8]. Recognition of objects, accidents, and behaviors in dynamic environments is still a great challenge in video surveillance [9], which should be carried out through objects detection, motion tracking and analyses, and understanding and recognition of other details with robust and efficient algorithms. Environments changes are so rich and varied that an online algorithm with universal significance is demanded towards the effective dangers detection and warning in dynamic environment changes [10–17].

Numerous algorithms have been developed to tackle video recognition challenges in various environments; however, a full understanding of environmental implications to video recognition efficiency demands learning models with universal significance (ignoring uncontrolled differences in real scenarios) [18–27]. That is the essential reason why the current online algorithms, even for latest algorithms, for
example, the latest models for tackling crowd segmentation for the high-dimensional, large-scale anomaly detection, still encounter considerable uncertainties [23, 24]. How to evaluate and compute the regulated attention in implications of the surrounding environment changes and, furthermore, how to define the warning level of EC-implications to video recognition efficiency should be research priorities towards precision security in intelligent surveillance [21–27].

It has been widely recognized that video surveillance should consider the implications of surrounding environments changes to video surveillance, but until now, there are still no models for a universal evaluation of EC-implications to video recognition efficiency [4, 12–27]. To solve the unresolved issues associated with uncontrolled EC-implications, various novel optimization models were proposed and further applied in current learning systems [13–15]. Robustness and efficiency of some online algorithms in tackling special EC-implications in special scenarios were validated in a series of previous studies although, until now, universal models for computation of the attention value and warning level of EC-implications to video recognition efficiency remain unaddressed and, hence, an emergent issue is improving the current surveillance systems [16, 17].

Objectives in this study are (1) to present a brain-inspired computing of warning level of the implications of surrounding environment changes to video recognition efficiency, (2) to model brain cognition processes and establish theoretical models for precision computation of attention value of EC, and (3) to highlight necessity of introducing proposed models in critical applications.

2. Preliminary Formulation

A conceptual framework of precision security to integrate video surveillance with EC is shown in Figure 1. Danger detection in EC-implications is of great complexity because of features diversity. Precision security aims to present a better understanding of EC-implications to danger detection efficiency in sensitive areas and allows us to consider not only “who are dangerous” but also “who are in danger” and to reduce uncertainties in uncontrolled and complicated real scenarios [28–31].

Brain cognition of EC-implications can be approached in four processes, data acquirement, classification, computation, and inference. Throughout the paper, the original, classified, computed, and inferred data, respectively, are denoted by $EC_1$, $EC_2$, $EC_3$, and $EC_4$. Obviously, $EC_2$ generates $EC_{i+1}$ ($i = 1, 2, 3$). To reduce uncertainty, assume that only $EC_3$ and $EC_4$ contribute to dispelling the EC-implications and generate regulated attention-effective data (denoted by $S$), which is generated from determination by $EC_4$ (with a contribution $\theta_3$) and a part of $EC_3$ (with a contribution $\theta_4$).
Complexity

Denote by $I^{(k)}$ amounts of newly generated effective data in the $k$-th brain learning period, $k = 1, 2, 3, \ldots$. Denote $A_i(t)$ as the amounts of EC$i$ at the $t$-th frame and let $A_{i0}^{(k)}, A_{i1}^{(k)}$ be EC$i$ at the beginning and end of the $k$-th period, respectively, $i = 1, 2, 3, 4, k = 1, 2, 3, \ldots$. Assume that the average efficiency of data exploitation is $r$ and employ a function $\delta(x) = 1/(1 + x)$ to estimate EC loss. Let $s_i$ be degree of EC$i$ importance and $\beta_i$ be the EC$i$ contributions to $S$, $i = 1, 2, 3, 4$; it is clear that $s_1 = s_2 = 0$, $\beta_1 = \beta_2 = 0$, $\beta_3 < \beta_4 = 1$. During the $k$-th learning period (with length $T_k$), define the theoretical quantification of attention value of EC-implications as the amounts of $S$ and define

\[
 f_k(t) = \begin{cases} 
 0, & \lambda_k \leq t \leq T, \\
 1, & 0 < t < \lambda_k, 
\end{cases}
\]

where $\lambda_k$ can be interpreted as EC attention-time ratio in the $k$-th learning period, $k = 1, 2, \ldots$.

Based on the performance of a rapid DL (deep learning) method, YOLO, which is one of the most efficient algorithms for objects detection, classification, and tracking [32–36], such implications of EC to video surveillance and the attention value and warning level are displayed in Figure 2.

Obviously, attention-time ratio of EC is reduced in regulated attention and EC-warning level (denoted by $\alpha$) is measured by corresponding time cost. Throughout the paper, computation of $\alpha$ is formulated as evaluation of time cost in implication-ratio changes from $\lambda_k$ to $\lambda_{k+1}$, $k = 1, 2, \ldots$.

3. Theoretical Analyses

Nonlinear functional analyses were confirmed suitable for solving the real scenario analyses and, exactly, multistage approach has been widely employed in simulating disaster responses [37–42]. But dangers of understanding and recognition in precision security are worthy of reconsideration to dispel EC-implications, utilizing determined EC-attention value and warning level for such implications. Recall that brain cognition of EC-implications can be theoretically approached in four processes and hence, correspondingly, the formulated problem should be resolved in a four-stage approach [40–46].

3.1. Attention Value of EC. Brain-inspired approach to attention value and warning level of EC are shown in Figure 3, where the EC-implications are manifested as an evolution of attention value and warning level. Such approach is independent of EC-types and hence it has universal significance. Regulated attention in brain-inspired data mining approach for behavior, accidents, and emotion understanding can be carried out through the whole video sampling, training, and recognition processes [47, 48].

First, we have

\[
 A_{i}^{(k)}(t) = A_{i0}^{(k)} e^{-rt}, \quad i = 1, 2, \\
 A_{j}^{(k)}(t) = A_{j0}^{(k)} e^{-r\lambda_k(\beta_{3} + \beta_{4})t}, \quad j = 3, 4,
\]

which imply

\[
 A_{11}^{(k)} = A_{10}^{(k)} e^{-r}, \quad A_{12}^{(k)} = A_{120}^{(k)} e^{-r}, \\
 A_{21}^{(k)} = A_{20}^{(k)} e^{-r}, \quad A_{22}^{(k)} = A_{220}^{(k)} e^{-r\lambda_k}, \\
 A_{31}^{(k)} = A_{30}^{(k)} e^{-r\lambda_k}, \quad A_{32}^{(k)} = A_{320}^{(k)} e^{-r\lambda_k}, \\
 A_{41}^{(k)} = A_{40}^{(k)} e^{-r\lambda_k}, \quad A_{42}^{(k)} = A_{420}^{(k)} e^{-r\lambda_k}.
\]
Suppose that EC\(_3\) can fully convert to EC\(_4\); we obtain

\[
A^{(k)}_{20} = A^{(k-1)}_{11},
\]

\[
A^{(k)}_{30} = A^{(k-1)}_{21},
\]

\[
A^{(k)}_{40} = A^{(k-1)}_{31} + A^{(k-1)}_{41},
\]

\[
A^{(k)}_{10} = \delta \left( I^{(k-1)} \right) j^{(k-1)}.
\]

Let \(\lambda_k = \lambda, k = 1, 2, \ldots\) From (2)–(4) and preliminary formulation, we have

\[
I^{(k-1)} = \int_0^\lambda \frac{A_3(t) + A_4(t)}{1-\lambda} \, dt = \frac{1-e^{-\lambda x \beta_3}}{(1-\lambda)(1-r+\beta_3)} A^{(k)}_{30} + \frac{1-e^{-\lambda x (r+\beta_4)}}{(1-\lambda)(1-r+\beta_4)} A^{(k)}_{40}.
\]

Therefore, theoretical quantification of \(S\) (i.e., the attention value of EC-implications) is

\[
S = \int_0^\lambda \left[ s_3 \beta_3 A_3(t) + s_4 \beta_4 A_4(t) \right] \, dt = \frac{s_3 \beta_3}{1-r+\beta_3} \left[ 1-e^{-\lambda x (r+\beta_3)} \right] A^{(k)}_{30} + \frac{s_4 \beta_4}{1-r+\beta_4} \left[ 1-e^{-\lambda x (r+\beta_4)} \right] A^{(k)}_{40} = s_3 \theta_3 + s_4 \theta_4.
\]

3.2. Determined Warning Level. It remains to determine warning level of EC-implications. To reduce time complexity of learning periods for the EC-universal significance, analyses can be divided into two cases: (1) time cost in different learning periods is independent or (2) considering periods are mutually dependent.

Within a single learning period, if EC evolution rate is fixed (denoted by \(x_r\)), then we have

\[
\frac{dx(t)}{dt} = \lambda(t) \cdot x_r.
\]
Complexity

Let \( \lambda(0) = \lambda_0 \); we have

\[
\lambda(t) = \lambda_0 \cdot e^{Kt}. 
\]  
(8)

Taking into account the variation of \( x_i \) within this period, for example, let \( x_i = k(K - \lambda(t)) \); we have

\[
\frac{d\lambda_i(t)}{dt} = \lambda_i(t) \cdot k_i \cdot (K - \lambda(t)),
\]
(9)

and hence

\[
\lambda_i(t) = \frac{\lambda_0 \cdot K}{\lambda_0 + (K - \lambda_0) \cdot e^{-k\lambda_0}},
\]
(10)

For a video with \( n \) learning periods, let \( x_{ij} = k_i(K_i - \lambda_i(t)) \); we have

\[
\frac{d\lambda_i(t)}{dt} = \lambda_i(t) \cdot k_i \cdot (K_i - \lambda_i(t)),
\]
(11)

\[ i = 1, 2, \ldots, n. \]

The solution of (11) is

\[
\lambda_i(t) = \frac{\lambda_0 \cdot K_i}{\lambda_0 + (K_i - \lambda_0) \cdot e^{-k\lambda_0}}, \quad i = 1, 2, \ldots, n. 
\]
(12)

Equivalently, we have

\[
(K_i - \lambda_0) \lambda_i(t) e^{-k\lambda_i} = (K_i - \lambda_i(t)) \lambda_{i0},
\]
(13)

\[ i = 1, 2, \ldots, n. \]

To simplify the representation of (13), define the following:

The Time-Parameters Matrices

\[
k = \text{diag}[k_1, k_2, \ldots, k_n],
\]
(14)

\[
K = \text{diag}[K_1, K_2, \ldots, K_n].
\]

The Original Status Matrix

\[
\lambda_0 = \text{diag}[\lambda_{i0}, \lambda_{20}, \ldots, \lambda_{n0}].
\]
(15)

The Dynamic Functions Matrix

\[
\lambda(t) = \text{diag}[\lambda_1(t), \lambda_2(t), \ldots, \lambda_n(t)].
\]
(16)

We obtain the matrix form of (13):

\[
(K - \lambda_0) \lambda(t) e^{-k\lambda_0} = (K - \lambda(t)) \lambda_0.
\]
(17)

Further considering relationship between surveillance videos, let \( (k_i, K_i) = (u_i(k_i), u_i(K_i)) \); then

\[
\lambda_i(t) = \frac{\lambda_0 \cdot u_i(K_i)}{\lambda_0 + (u_i(K_i) - \lambda_0) \cdot e^{-u_i(k_i)u_i(K_i)t}},
\]
(18)

\[ u_i = 1, \; i = 1, 2, \ldots, n. \]

The symmetric form of (17) is

\[
(u_i(K_i) - \lambda_0) \lambda_i(t) e^{-u_i(k_i)u_i(K_i)t} = (u_i(K_i) - \lambda_i(t)) \lambda_{i0}, \quad i = 1, 2, \ldots, n.
\]
(19)

Defining \( u = \text{diag}[u_1, u_2, \ldots, u_n] \), \( \lambda_0 = \text{diag}[^{\lambda_{10}} \lambda_{20}, \ldots, \lambda_{n0}] \), \( \lambda(t) = \text{diag}[\lambda_1(t), \lambda_2(t), \ldots, \lambda_n(t)] \), we obtain

\[
(u(K_i) - \lambda_0) \lambda(t) e^{-u(k_i)u(K_i)t} = (u(K_i) - \lambda(t)) \lambda_0,
\]
(20)

where \( u_i \) is the correlation function of the \( i \)-th video in the consider security system, \( i = 1, 2, \ldots, n \).

Finally, EC-warning level can be computed as time cost from \( \lambda_j \) to \( \lambda_{j+1} \), \( j = 1, 2, \ldots \). Regulated attention can be theoretically implemented in multidata fusion, learning, and modelling. Region of interest (ROI) or pedestrians of interest (POI) corresponds to GIS-data, including time, place, and EC through Internet of things applicable for real scenarios, as seen in Figure 4. It is worth noting that the 3D stereo generated from a 2D video sequence is advantageous to highlight EC evolution and therefore is also advantageous to determine length of learning periods.

4. Simulation and Discussion

Our proposed models in the present study are learning models with universal significance (ignoring uncontrolled differences in real scenarios), which aim to establish theoretical framework of the environmental implications to video recognition efficiency. It will serve for a universal evaluation of EC-implications to video recognition efficiency. Numerous algorithms have been developed to tackle video recognition challenges in various environments, but it is still difficult to describe the time complexity of learning periods. This can be largely attributed to the complexity of video recognition issues. Even for a given issue, it is not easy to determine learning periods for different EC-scenarios. Generally, attention-time ratio of EC is reduced in regulated attention and EC-warning level can be measured by corresponding time cost in reducing the attention-time ratio of EC. So we formulate the parameter \( \alpha \) as the time cost in implication-ratio changes from \( \lambda_k \) to \( \lambda_{k+1} \), \( k = 1, 2, \ldots \). For detailed analysis on the time complexity, some examples of learning periods for video detection and tracking in different surveillance scenarios are presented in Figure 5. One possible solution to treat the time complexity is to imbed proposed models into online algorithms in critical applications, utilizing these newly added examples and evidences.

Because of time complexity of learning periods, we give EC-attention values for simulation, ten videos with given EC-attention values in Table 1. Equations (17)–(20) are employed to simulate brain-inspired computing of corresponding EC-warning level.

Ignoring the association among ten surveillance videos, from (17) and (18), the EC-warning levels from \( \lambda_i \) to \( \lambda_{i+1} \) are \( \alpha_1 = 0.8868 \), \( \alpha_2 = 0.1363 \), \( \alpha_3 = 1.5691 \), and \( \alpha_4 = 0.9220 \), respectively, \( i = 1, 2, 3, 4 \). Taking into account the association among ten surveillance videos, utilizing (19) and (20) and letting \( k_1 = 0.4503 \), \( K_2 = -1.6369 \) and finding a suitable
Table 1: Attention values of ten videos with unneglectable EC.

<table>
<thead>
<tr>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
<th>Period 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>30.21</td>
<td>2.68</td>
<td>1.01</td>
<td>0.81</td>
<td>0.33</td>
</tr>
<tr>
<td>29.88</td>
<td>1.62</td>
<td>1.43</td>
<td>0.76</td>
<td>0.31</td>
</tr>
<tr>
<td>26.66</td>
<td>1.23</td>
<td>1.29</td>
<td>0.71</td>
<td>0.28</td>
</tr>
<tr>
<td>25.12</td>
<td>1.28</td>
<td>1.26</td>
<td>0.65</td>
<td>0.27</td>
</tr>
<tr>
<td>23.33</td>
<td>1.24</td>
<td>1.11</td>
<td>0.54</td>
<td>0.26</td>
</tr>
<tr>
<td>13.33</td>
<td>1.18</td>
<td>1.12</td>
<td>0.48</td>
<td>0.21</td>
</tr>
<tr>
<td>0.32</td>
<td>1.02</td>
<td>1.01</td>
<td>0.39</td>
<td>0.22</td>
</tr>
<tr>
<td>0.18</td>
<td>0.98</td>
<td>0.96</td>
<td>0.35</td>
<td>0.23</td>
</tr>
<tr>
<td>0.17</td>
<td>0.86</td>
<td>0.91</td>
<td>0.31</td>
<td>0.23</td>
</tr>
<tr>
<td>0.16</td>
<td>0.79</td>
<td>0.68</td>
<td>0.27</td>
<td>0.25</td>
</tr>
</tbody>
</table>

association function (here \( u(x) = ax + blnx \)), EC-warning levels from \( \lambda_i \) to \( \lambda_{i+1} \) are \( \alpha_1 = 0.4096, \alpha_2 = 0.0984, \alpha_3 = 0.6314, \) and \( \alpha_4 = 0.9220 \), respectively, \( i = 1, 2, 3, 4 \).

Characterizing EC-warning level and the implied dangers is helpful for learning how well can potential dangers be detected by video surveillance in changing environments, especially in unmanned driving, where one major bottleneck is finding effective and efficient algorithms for the danger detection and caution, majorly due to lack of adaptive attention in utilized learning systems [49–51]. Numerous issues remain unresolved, a part of which are resulted from poorly understood EC-implications [52–58]. Brain-inspired modelling approach to such implications in the present study majorly depends on amounts of attention data and length of attention time, ignoring the differences in real scenarios. Therefore, the proposed models have universal significance for its critical applications. It is therefore necessary to consider integration of proposed models with the online surveillance algorithms towards precision security [59–61]. Such precision security can be a great challenge because that performance degradation of video recognition efficiency in critical environments has been demonstrated in some previous studies [6, 17, 21, 35].

For special scenarios when EC-implications are not significant, integration of our models with online algorithms is not necessary. Computation can be largely simplified in special applications. Taking the lane detection as an example, the biological principles are to detect and recognize a line, which can work well even if the lanes are partly missing [62–64], as seen in Figure 6.

For complex applications, however, imbedding proposed models in current security systems becomes necessary, such as compressive sensing for sparse tracking [18] (it can be improved as locally compressive sensing within ROI), VIBE algorithm for real-time object detection from a moving camera [19], Adaboost algorithm for noise-detection in ROI [20], optical flow for robots’ recognition of environments [21], SVM clustering for accidents classification [22], deep learning algorithms for anomaly detection, crow analysis, and hierarchical tracking within ROI [23–27]. Objects understanding and detection in dynamic environment changes are usually
based on the adaptive background subtraction and other objects recognition methods [17, 21, 35, 65–68]. A preliminary scheme for the practical integration of proposed models with these algorithms is presented in Figure 7, where smog as a global environmental change has significant implications to video behaviors recognition and loitering detection within a hovering period of two persons; only half of hovering behaviors is detected; only one person is red-highlighted and the other person is always in a green rectangle, indicating the degradation of video surveillance efficiency within the considered periods under any real challenging scenarios. It is worth noting that the proposed models have analytic solutions and the time cost in each iteration is much shorter than the time cost of any video recognition algorithms. Therefore, imbedding the proposed models in current security systems for critical applications is not only necessary but also feasible; proposed models can work well with any online algorithms without a great loss in surveillance efficiency.

5. Conclusion

Despite previous studies on algorithms for video surveillance in various environments, there are still considerable uncertainties in objects detection, classification, and tracking. Understanding and recognition of implications of surrounding environment changes to surveillance efficiency are still very limited. Brain-inspired modelling approach to such implications in the present study majorly depends on the amounts of attention data and attention time, ignoring difference in real scenarios. Therefore, proposed models represent biological principles of computational intelligence and have universal significance for its practical integration with online algorithms. Nevertheless, a full understanding of complexity of learning periods for different EC-scenarios is still necessary. This is also a next research priority towards a universal evaluation of implications of the surrounding environments changes to video recognition efficiency.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was financially supported by the Shenzhen Basic Research Project (JCYJ20150630114942260), the

Figure 5: Examples of learning periods for video recognition in different surveillance scenarios for detailed analysis on the time complexity, under implications of smog (in the blue-rectangles, captured by an android camera), sandstorm (in the green rectangles, captured by a webcam on mobile), blizzard (in those yellow rectangles with videos collected from a drone), and truck exhaust (in the red rectangles with videos collected by a driving recorder).
Figure 6: An example does not normally require integration with proposed models; truck exhaust as a regional environmental change has no significant implications to efficiency of the lanes detection and warning of conflict danger. Lanes detection is always working well although a part of detected lane is temporally missing (highlighted by the red caution-texts "Right Departure") within the period of right departures and a robust and efficient warning of conflict danger (highlighted by the red caution-texts "Conflict Alert") works well simultaneously during the period (Frame #156 to #164).

Figure 7: An example demands integration of proposed models with other online algorithms; smog as a global environmental change has significant implications to loitering detection within a hovering period of two persons, where only half warning of hovering behaviors is detected (highlighted by the red rectangles). A preliminary scheme for practical integration is subsequently presented.
National Natural Science Foundation of China (41571299), the CAS “Light of West China” Program (XBB5-2014-16), and the “Thousand Talents” plan (Y474161).

References


C. J. Littles, S. S. Pilyugin, and T. K. Frazer, “A combined inverse
J. H. Ruan, X. P. Wang, and Y. Shi, “A two-stage approach for medical supplies intermodal transportation in large-scale disaster responses,”
Submit your manuscripts at www.hindawi.com