

Event Analogy Based Privacy Preservation in Visual Surveillance

Wei Qi Yan^{1,2(✉)} and Feng Liu²

¹ Auckland University of Technology, Auckland, New Zealand
weiqi.yan@aut.ac.nz

² Chinese Academy of Sciences, Beijing, China

Abstract. Privacy preservation as a thorny problem in surveillance has been arisen because of its relevance to human right, however it has not been completely solved yet today. In this paper, we investigate this existing problem and expect to get ride of those intuitive methods such as pixelization, blurring or mosaicking on human face regions through object tracking. We detail privacy preservation at event level and thereafter choose suitable events represented by motion pictures in virtual reality to replace those events of surveillance in real reality. The advantage of taking use of this approach is to leverage the utility and privacy of surveillance events. The outcome will not affect visual effects and surveillance security much however it is able to achieve the objectives of privacy preservation.

Keywords: Event analogy · Surveillance event · Privacy preservation · Real reality · Virtual reality

1 Introduction

Surveillance has been accepted as an effective way to protect our community especially after the September 11 attacks. Nowadays digital cameras are being networked at every corner of a metropolitan and monitoring our actions and behaviours in real time. This has made our ordinary lives too secure to have our own privacy. At anytime from anywhere we feel that a hidden eye is looking at the earth dwellers. Therefore, the emerging problem is how to protect our privacy especially in the era when it has high risk that surveillance data is abusively utilized.

The traditional ways to preserve privacy in visual surveillance are mosaicking, pixelization or scrambling human face regions [18] shown as Fig. 1. But apparently this is not enough since from the acquired clothes, behaviours or gait, even from a contour, silhouette or blob, several dots, we still are able to discern who this person is as shown in Fig. 2. Even if a face is not clearly seen from very far distance, such as a basketball or soccer player, we are still able to infer who the person is. Particularly from those processed images, photos or cartoon motion pictures, the similarity is still existential. This rolls out the motivation



Fig. 1. Privacy preservation using: (a) mosaicking (b) Gaussian blurring (c) scrambling

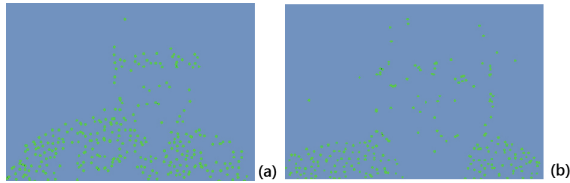


Fig. 2. An example of a perceptual human face in dots

that the best solution of privacy preservation is to completely remove the privacy information from the video frames. However the commitment will undoubtedly diminish the utility and visibility of surveillance videos.

Utility refers to video usage for various purposes. When an incident happens, we need track back and search for the persons and objects related to the incident. While the conventional ways of privacy protection such as image mosaicking, blurring and pixelization easily provoke the content damaging, if the surveillance video frames have been completely obstructed, that is equivalent to acclaim that this video has to be casted aside. Thus, the situation requires us find a way to leverage the utility, visibility and privacy of surveillance videos. Our goal of this paper is to resolve this tough problem.

In this paper, our idea is to replace a surveillance event in real reality using a resembling event presented by motion pictures in virtual reality, the two picture sequences carry similar semantic events however the privacy information in one of them has gone. Therefore, we segment our surveillance events into several states. For each state, we find the pictures which could be modelled and replaced by motion pictures. Aftermaths we still could understand the event content, however we have thrown away the annoying human privacy information.

We justify our idea as an effective way for preserving privacy. For an instance, in a typically monitored corridor, we use a walking Mickey Mouse to substitute a man for displaying purpose who is walking through from left to right or from right to left, the man may perambulate to pass this site, thereafter the mouse will be viewed in the correspondingly sluggish way such as entering, walking, standing, existing, alarming etc. If it indeed has an incident, namely the alarming state is activated, only the authorized security staff has the privilege to review the surveillance events, but normally this analogy based replacement for the purpose of privacy preservation is much reasonable for catering to unauthorized viewers.

The challenge of this research work is to seek the matching between two events presented by two groups of motion pictures using analogy. We therefore call this analogy at event level as event analogy. We have successfully developed a concept of video analogy based on image analogy [5,20], however event analogy gets out of the box at physical or object level and aims at semantics. In this paper, we will select suitable events in virtual reality to replace the surveillance events in real reality for the sake of privacy preservation.

Our idea was inspired by a movie pertained to bus surveillance. In the 1994 Hollywood movie “The Speed”, a young cop must prevent a bomb exploding aboard a city bus by keeping its speed above 50 mph. The LAPD interrupted the live broadcasting connected to the on-board bus video surveillance system for the public and replayed a pre-recorded cassette video only having a few frames shown as Fig. 3, however the rival could not make out this minor change timely so that the bus passengers have ample time to alight and get rescued successfully. This story implies that event analogy could save human lives including privacy in very exigent time.



Fig. 3. The replaced picture of 1994 movie: The Speed

Event is defined as a semantic unit which bridges the gap between semantic world and cyberspace [19]. An event has the basic components such as who(object), when(time stamp), where(site), what(description), and why(reasoning). As a fundamental structure, discrete events could be stored in computers as logs for the purposes of analysis and archiving.

Our goal in this paper is to leverage the utility and security of surveillance videos so as to preserve human privacy in surveillance. The rest of this paper is organized as follow. The related work will be introduced in Sect. 2, our contributions will be presented in Sect. 3, Sect. 4 will provide the experimental results and analysis, conclusion and future work will be addressed in Sect. 5.

2 Related Work

Analogy as said is “The art of the metaphor” [8,9]. Metaphor is a rhetoric which has often been applied to our oral and writing presentations. It’s believable that we always explain a profound and abstractive theory using an akin

easy-understanding story to feed our audience. The concept analogy was from cognition science [8] and have been digitalized as a reasoning or inference method in Artificial Intelligence(AI).

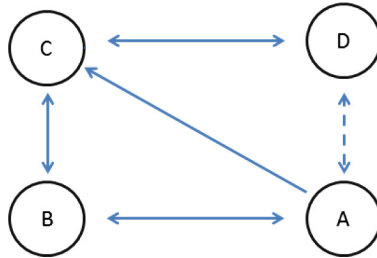


Fig. 4. The mechanism of event analogy

Figure 4 shows the fundamental relationships amongst participants of an analogy. Suppose we have similar events A and B as our start point, C and B are similar but C has its outperforming attribute such as with visibility without privacy. We envision transferring the unique attribute of C to the event D , where D resembles from A . Therefore, we see the analogy operation as a kind of fundamental reasoning based on the facts at hand to get the unknown knowledge. The intuitive explanation of an analogy is that if event A could remove its privacy, then the privacy in event B also could be removed. However the visibility of two events is still preserved. In a nutshell, we denote an analogy mathematically as: if $A \Leftrightarrow B$, $B \propto C$ and $A \propto D$, then $C \Leftrightarrow D$.

Analogy has been applied to curves and geometry objects initially [6, 7]. The concept image analogy is a metaphor between two digital images [5] which has been applied to render a gray scale image using another color image. Albeit we do not exactly affirm the colors of the photo, we still could map the colors of this scene of today to the gray scale image using color transferring technologies based on texture synthesis [1, 3].

Video analogy was derived from image analogy [4, 19]. Assume we have two similar videos at hand, we therefore create a relationship and bridge the gap between two videos. Thus, we could transfer some attributes of one video to the other which is lack of this attribute such as color, motion, contrast, etc. Based on the merit of video analogy, the media aesthetics could be transferred to amateur's craft work which has the longing to be forged as an art masterpiece, etc [10–14].

In multimedia analysis, the concept event analogy is created at semantic level which differs from object analogy, e.g. image analogy and video analogy both are manipulated at physical level. A semantic event could be presented in both real reality and visual reality, therefore a semantic event could have or have not privacy. Thereafter through attribute transferring of event analogy, we have the opportunity to add or remove privacy information from one event by analogizing

the other event meanwhile keeping the semantic meaning. In this paper, we will work for the theory and implementation of event analogy.

Privacy of surveillance video [17] has been modelled by the parameters ‘who’, ‘when’ and ‘where’ due to the applications of events. The detected pedestrian face and head in a surveillance video usually are obscured by encrypting for the purpose of privacy preservation [18]. A privacy preservation method adopts data transformation involving the use of selective obfuscation and global operations to provide robust privacy [15].

Conventional privacy protection methods directly consider explicit privacy losing (such as facial information) and ignore other implicit channels. A privacy model [16] consolidates the identity leakage through both implicit and explicit channels. The computational model using a combination of quantisation and blurring also provides the best tradeoff between privacy and utility.

Unlike those existing work, the focus of this paper is on preserving privacy existing in surveillance events. The novelty of this paper is that it is the first time to create the concept event analogy in which we adopt the event in virtual reality to replace the surveillance event happened in real reality while conveying the same semantics. The replacement will remove privacy information in a surveillance event so as to leverage the utility and security of surveillance events.

3 Our Contributions

To the best of our knowledge, privacy preservation using event analogy is a brand-new approach. However, the main challenge is how to find the resembling event presented by motion pictures to replace the surveillance events in real reality. Therefore the first problem is how to optimize the motion pictures and remove the privacy information to match the surveillance video in real reality. Thus time line from the surveillance video has to be followed, correspondingly the motion pictures should be put on the time line flexibly, this is similar to achieve the results of synthesizing a multimedia message [21].

3.1 Surveillance Events

In surveillance environment, usually cameras will be deployed at a fixed site, motion pictures captured by a camera will show the events having steady patterns though the cameras have the functionalities such as panning, tilting and zooming. After thoroughly observed these events, we find in indoor environment a walker usually toddles from left to right or from right to left within a framed route such as corridor or walkway. While in outdoor environment the cameras are usually operating from morning to night under all weather conditions, the objects encapsulate moving vehicles and pedestrians restrained in their own track rigorously.

In this paper, we capture surveillance events using Finite State Machine (FSM) shown in Fig.5. In the scenarios of walking through a corridor,

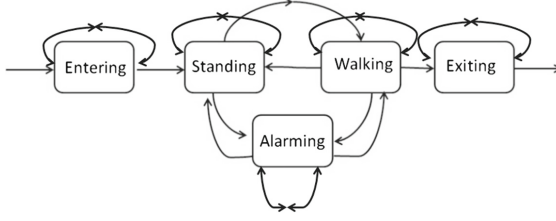


Fig. 5. Surveillance event capturing using FSM

we set 5 states including alarming. Our surveillance event capturing is based on the state changes [19]. The pseudo code for FSM based event capturing is shown as below algorithm.

Algorithm. FSM based surveillance event capturing

```

Input: Image sequence
Output: Captured events: e = {W(Walking), A(Alarming), S(Standing)}
Line 1. Initialization: set flags f := 0 and t := 0;
Line 2. while (True){
Line 3. Detect state: t;
Line 4. If (t == 0) && (f == 0) return;
Line 5. If (t == W|A|S) && (f == 0) {f := t; }
Line 6. If (t == W|A|S) && (f == t) return;
Line 7. If (t == W|A|S) && (f != t) {e := f; f := t; output: e;}
Line 8. If (t == 0) && (f != 0) {e := f; f := t; output: e;}
Line 9. }

```

In the event of detection of surveillance events, state changes are usually detected based on local intensity histogram (N_x^l, N_y^l, N_t^l) from spatial-temporal viewpoint, motion changes $\triangle I = (I_x, I_y, I_t) = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t})$ will be normalized so as to feed the distance calculator based on χ^2 -divergence [23–25],

$$d(H_1, H_2) = \sqrt{\frac{1}{3L} \sum_{l,i,k} \frac{|h_{1k}^l(i) - h_{2k}^l(i)|^2}{h_{1k}^l(i) + h_{2k}^l(i)}} \quad (1)$$

where an action is represented by a set of nine one-dimensional histograms: $\{h_x^1, h_y^1, h_t^1, h_x^2, h_y^2, h_t^2, h_x^3, h_y^3, h_t^3\}$, B is the histogram bin numbers of each video frame, L is total frame number of an image sequence.

From our observations, we find that surveillance events calculated by Eq. (1) have their own patterns owning the merits such as discriminative and covering. We therefore have the opportunity to seek the typical motion pictures with a specific pattern, such as the cartoon GIF pictures which could be played iteratively and are suitable for presenting these surveillance events. Therefore, adjustment of these motion pictures is entailed to match the necessity of surveillance events.

3.2 Event Analogy

Event analogy is derived from cognition sciences in AI which has been digitalized in curve analogy of geometry [6], image analogy in computer graphics [5] and video analogy in multimedia analysis and synthesis [20]. In visual surveillance, event analogy is reckoned to be applied to privacy preservation in Fig. 4. Hence we define event analogy as the below Definition 1.

Definition 1 (Event Analogy). If $\forall e \in \{e_A, e_B, e_C, e_D\}$, $e_A \Leftrightarrow e_B$, $e_A \propto e_D$, $e_B \propto e_C$, then $e_C \Leftrightarrow e_D$.

Following Definition 1, the probability which event e_D may happen could be predicted by using Dynamic Bayesian Network (DBN) as a directed graph in Eq. (2),

$$p(e_D) = p(e_D|e_C) \cdot p(e_D|e_A) = p(e_C|e_B) \cdot p(e_B)p(e_C|e_A) \cdot p(e_A) \cdot p(e_D|e_A) \quad (2)$$

Since $p(e_B) = 1$, $p(e_A) = 1$, thus,

$$p(e_D) = p(e_C|e_B) \cdot p(e_C|e_A) \cdot p(e_D|e_A) \quad (3)$$

This simplification reveals that whether the event e_D will be happened or not, it is mostly decided by the relationship between e_C and e_B , e_D and e_A since e_A and e_B have been given as the known condition.

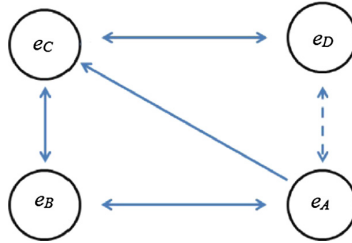


Fig. 6. Event analogy

Equation (3) reflects the ground truth of event analogy. We presume that event e_C has the state set $S_C = \{s_C^1, s_C^2, \dots, s_C^n\} \subseteq S_B$ meanwhile for the event e_A and event e_D we have the relationship $S_D = \{s_D^1, s_D^2, \dots, s_D^n\} \subseteq S_A$.

In this paper, we anticipate the overlapping could correctly reflect visibility of the event however its privacy will be removed. Figure 7 is an example of event analogy, we used the video provided in the surveillance data set: CAVIAR to demonstrate a walker passing through a shop in a mall. The state diagram with video frames depicts the typical events of a walker when passing through a monitored corridor: *entering*, *standing*, *passing*, *alarming*, and *exiting*. The states could be switched between each other due to changes of the guard condition

and actions. In order to analogize the event and remove the privacy information, we find an animal cartoon from online GIF picture store which has the similar state changes. Namely, we detect the state changes, we find cartoon pictures presenting the similar states, finally the privacy region on the surveillance video frames has been overlapped and the privacy of the event has been removed (Fig. 6).

State diagram in Fig. 7 illustrates the connections between the events, states and surveillance video frames. This example epulides how we could leverage human privacy in a surveillance event using event analogy.

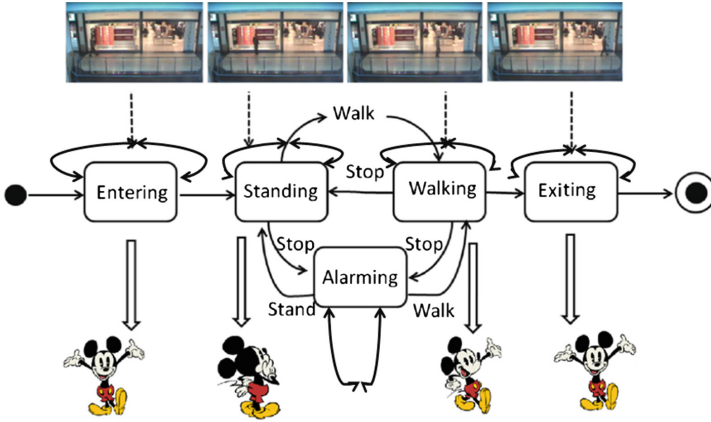


Fig. 7. A state diagram and video frames show an instance of event analogy for privacy preservation in visual surveillance

4 Analysis

We implement our privacy preservation of surveillance events using event analogy. Shown as Figs. 8, 10, 11 and 12, we detect moving object, track the object and find the state changes of an event in a surveillance scenario. In Figs. 8 and 11, we detect the ‘entering’, ‘standing still’ and ‘exit’ states of the surveillance event, therefore we could cover the moving object using cartoon characters.

In Fig. 9, we find the cartoon pictures from public web sites with swinging the right-hand, swinging the left-hand and standing still in virtual reality, the six cartoons represent the states of two opposite walking directions: left to right and right to left through the corridor, the actions of cartoon characters could represent the states of surveillance events in real reality.

The differences of surveillance videos before and after moving object overlapping by cartoon characters are measured by histogram based image entropy. In another word, the differences between them are approaching to the privacy difference.

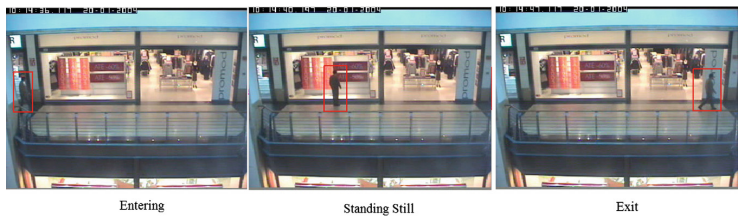


Fig. 8. Object tracking from left to right with standing state

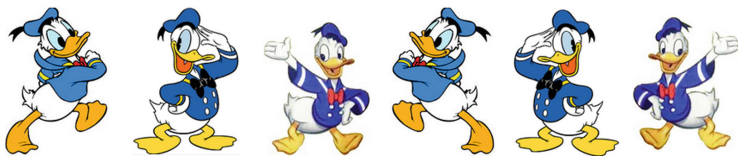


Fig. 9. The Donald Duck’s actions in virtual reality for representing surveillance events in reality



Fig. 10. Overlapping moving object from left to right

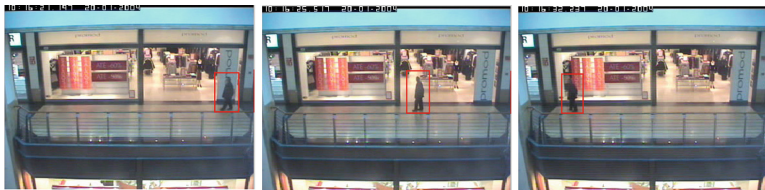


Fig. 11. Moving object tracking from right to left

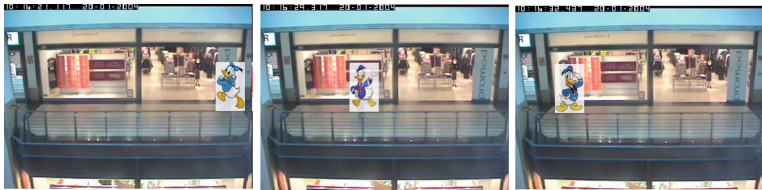


Fig. 12. Overlapping moving object from right to left

From the two results shown as Figs. 13 and 14, we see that the videos overlapped by cartoon characters have much entropy than that of original ones. This is due to the image region overlapped by cartoon characters has much information than the original. However after the overlapping operation, the privacy intensity of the surveillance video has gone. The viewers could not find any privacy information related to the moving object from the processed videos. Thus it achieves our goal of privacy preservation of this paper.

Therefore, we have the opportunity to choose the best event presented by motion pictures in virtual reality. Using event analogy, we could find the

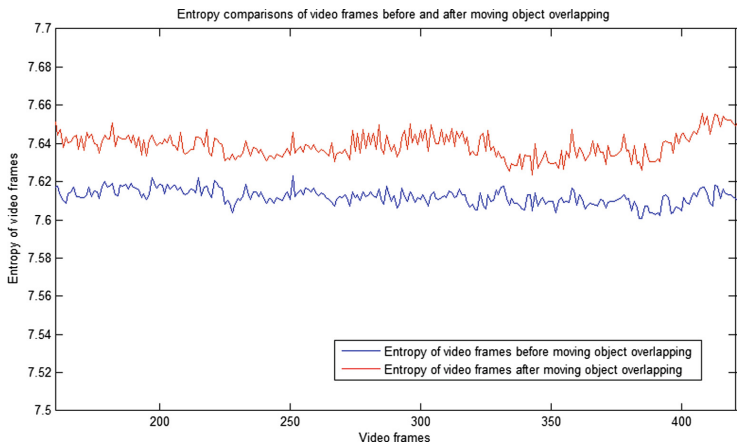


Fig. 13. Entropy comparisons of surveillance video 1 before and after moving object overlapping by cartoon characters

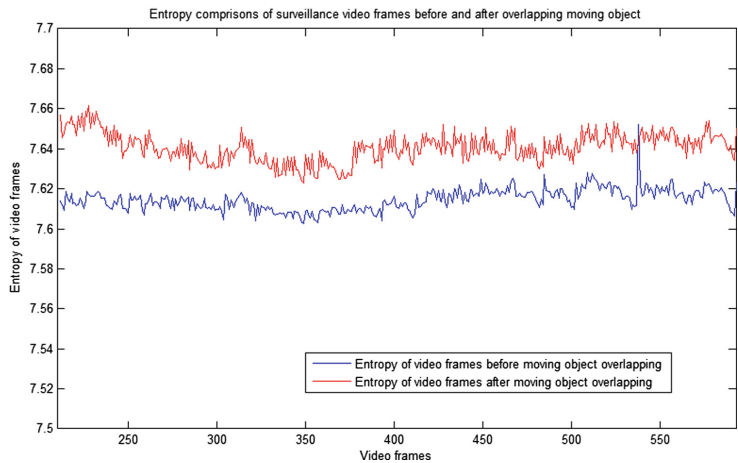


Fig. 14. Entropy comparisons of surveillance video 2 before and after moving object overlapping by cartoon characters

pertinent cartoon pictures in virtual reality to replace the events in real reality. Therefore we have to acquire event first, then preserve privacy, this is much different from those privacy preservation directly using blurring, mosaicking and blurring, thus the technical advance requirement is very high.

5 Conclusion

In this paper, we leverage utility and privacy of surveillance videos using event analogy. Our core idea is to overlap human privacy region of surveillance motion pictures using selected animated cartoons so as to preserve human privacy. It's the first time that we are in use of this concept: event analogy to seek the similarity in virtual reality and real reality of surveillance events. In future, we will embark on privacy preservation of visual surveillance and seek the best form in presenting surveillance events.

References

1. Reinhard, E., Ashikhmin, M., Gooch, B., Shirley, P.: Color transfer between images. *IEEE Comput. Graph. Appl.* **21**(5), 34–41 (2001)
2. Klette, R.: *Concise Computer Vision*. Springer, London (2014)
3. Welsh, T., Ashikhmin, M., Mueller, K.: Transferring color to greyscale images. In: *Computer graphics and interactive techniques*, pp. 277–280, USA (2002)
4. Yan, W., Kankanhalli, M.: Colorizing infrared home videos. In: *IEEE ICME 2003*, pp. 97–100, USA (2003)
5. Hertzmann, A., Jacobs, C., Oliver, N., Curless, B., Salesin, D.: Image analogies. In: *Computer graphics and interactive techniques*, pp. 327–340, USA (2001)
6. Evans, T.: A program for the solution of geometric analogy intelligence test questions. In: *Semantic Information Processing*. MIT Press, New York (1968)
7. Hertzmann, A., Oliver, N., Curless, B., Seitz, S.: Curve analogies. In: *Eurographics workshop on Rendering*, Italy (2002)
8. Winston, P.: Learning and reasoning by analogy. *Commun. ACM* **23**(12), 689–703 (1980)
9. Gentner, D.: Structure mapping: a theoretical framework for analogy. *Cogn. Sci.* **7**(2), 155–170 (1983)
10. Dorai, C., Venkatesh, S.: Bridging the semantic gap with computational media aesthetics. *IEEE MultiMed.* **10**(2), 15–17 (2003)
11. William, T., Touis, R., Egon, C.: Example-based super-resolution. *IEEE Comput. Graph. Appl.* **22**, 56–65 (2002)
12. Adams, B., Dorai, C., Venkatesh, S.: Automated film rhythm extraction for scene analysis. In: *Proceedings of the IEEE ICME 2001*, pp. 1192–1195, Japan (2001)
13. Adams, B., Venkatesh, S., Dorai, C.: Finding the beat: an analysis of the rhythmic elements of motion pictures. *Int. J. Image Graph.* **2**(2), 215–245 (2002)
14. Herbert, Z.: *Sight, Sound, Motion: Applied Media Aesthetics*. Wadsworth Publishing Company, Belmont (1999)
15. Saini, M., Atrey, P., Mehrotra, S.: Adaptive transformation for robust privacy protection in video surveillance. *Int. J. Adv. Multimed.* **2012**, 4 (2012)
16. Saini, M., Atrey, P., Mehrotra, S., Kankanhalli, M.: Privacy aware publication of surveillance video. *Int. J. Trust Manage. Comput. Commun.* **1**(1), 23–51 (2012)

17. Saini, M., Atrey, P., Mehrotra, S., Emmanuel, S., Kankanhalli, M.: Privacy modeling for video data publication. In: IEEE ICME 2010, Singapore (2010)
18. Zhang, P., Thomas, T., Emmanuel, S., Kankanhalli, M.: Privacy preserving video surveillance using pedestrian tracking mechanism. In: ACM Workshop on Multimedia in Forensics, Security and Intelligence (2010)
19. Yan, W., Kieran, D., Rafatirad, S., Jain, R.: A comprehensive study of visual event computing. *Multimed. Tools Appl.* **55**(3), 443–481 (2011)
20. Yan, W., Kankanhalli, M.: Analogies based video editing. *Multimed. Syst.* **11**(1), 3–18 (2005)
21. Yan, W., Kankanhalli, M.: Multimedia simplification for optimized MMS synthesis. *ACM Trans. Multimed. Comput. Commun. Appl.* **3**(1) (2007)
22. Rogez, G., Rihan, J., Orrite, C., Torr, P.: Fast human pose detection using randomized hierarchical cascades of rejectors. *Int. J. Comput. Vis.* **99**(1), 25–52 (2012)
23. Prest, A., Ferrari, V., Schmid, C.: Explicit modeling of human-object interactions in realistic videos. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(4), 835–848 (2013)
24. Zelnik-Manor, L., Irani, M.: Event-Based Video Analysis. In: IEEE CVPR (2001)
25. Zelnik-Manor, L., Irani, M.: Statistical analysis of dynamic actions. *IEEE PAMI* **28**(9), 1530–1535 (2006)