Expert Systems with Applications 42 (2015) 4177–4195

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Review

Visual privacy protection methods: A survey

José Ramón Padilla-López [a,*], Alexandros Andre Chaaraoui [a], Francisco Flórez-Revuelta [b]

[a] Department of Computer Technology, University of Alicante, P.O. Box 99, E-03080 Alicante, Spain
[b] Faculty of Science, Engineering and Computing, Kingston University, Penrhyn Road, KT1 2EE, Kingston upon Thames, United Kingdom

ARTICLE INFO

Article history:
Available online 31 January 2015

Keywords:
Visual privacy protection
Video surveillance
Ambient-assisted living
Computer vision
Image processing

ABSTRACT

Recent advances in computer vision technologies have made possible the development of intelligent monitoring systems for video surveillance and ambient-assisted living. By using this technology, these systems are able to automatically interpret visual data from the environment and perform tasks that would have been unthinkable years ago. These achievements represent a radical improvement but they also suppose a new threat to individual's privacy. The new capabilities of such systems give them the ability to collect and index a huge amount of private information about each individual. Next-generation systems have to solve this issue in order to obtain the users' acceptance. Therefore, there is a need for mechanisms or tools to protect and preserve people's privacy. This paper seeks to clarify how privacy can be protected in imagery data, so as a main contribution a comprehensive classification of the protection methods for visual privacy as well as an up-to-date review of them are provided. A survey of the existing privacy-aware intelligent monitoring systems and a valuable discussion of important aspects of visual privacy are also provided.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

It can be observed that world population is ageing. In fact, it is estimated that population over 50 will rise by 35% between 2005 and 2050, and those over 85 will triple by 2050 (EC, 2010). Furthermore, the number of people in long-term care living alone is expected to increase by 74% in Japan, 54% in France and 41% in the US (EC, 2008). Therefore, this situation will not be sustainable in the near future, unless new solutions for the support of the older people, which take into account their needs, are developed.

Ambient-assisted living (AAL) aims to provide a solution to this situation. AAL applications use information and communication technologies to provide support to people so as to increase their autonomy and well-being. Video cameras are being used more and more frequently in AAL applications because they allow to get rich visual information from the environment. The advances produced in the last decades have contributed to this. The computational power has been increased while, at the same time, costs have been reduced. Furthermore, computer vision advances have given video cameras the ability of 'seeing', becoming smart cameras (Fleck & Strasser, 2008). This has enabled the development of vision-based intelligent monitoring systems that are able to automatically extract useful information from visual data to analyse actions, activities and behaviours (Chaaraoui, Climent-Pérez, & Flórez-Revuelta, 2012b), both for individuals and crowds, monitoring, recording and indexing video bitstreams (Tian et al., 2008). By installing networks of cameras in people homes or care homes, novel vision-based telecare services are being developed in order to support the older and disabled people (Cardinaux, Bhowmik, Abhayaratne, & Hawley, 2011). But these new technologies also suppose a new threat to individual's privacy.

Traditionally, cameras are used in public spaces for surveillance services in streets, parking lots, banks, airports, train stations, shopping centres, museums, sports installations and many others. It is estimated that there is an average of one camera for every 32 citizens in the UK, one of the most camera-covered countries in the world (Gerrard & Thompson, 2011). In short, video cameras are mainly used in outdoor environments and in public places, but they are not commonly used within private environments due to people's concerns about privacy. Generally, the use of video cameras in public places has been tolerated or accepted by citizens, whereas their use in private spaces has been refused. There may be several reasons to explain this difference. On the one hand, the perceived public-safety benefits favour the usage of cameras in public places for crime prevention, fight against terrorism and others. On the other hand, there is a widespread belief that while staying in public environments, people's sensitive information will not be exposed. Finally, there are some attitudes which have also contributed to accept their use, for example, to assume that anyone demanding privacy must have something to hide (Caloyannides, 2003).

* Corresponding author.
E-mail addresses: jpadilla@dtic.ua.es (J.R. Padilla-López), alexandros@dtic.ua.es (A.A. Chaaraoui), F.Florez@kingston.ac.uk (F. Flórez-Revuelta).

http://dx.doi.org/10.1016/j.eswa.2015.01.041
0957-4174/© 2015 Elsevier Ltd. All rights reserved.

In traditional video surveillance systems cameras are managed by human operators that constantly monitor the screens searching for specific activities or incidents. As estimated by Dee and Velastin (2008), the ratio between human operators and screens is around 16 displays for every operator in four local authority installations within the UK. Although they can only really watch 1–4 screens at once (Wallace & Diffley, 1988), this does not prevent abuses of these systems by their operators. Furthermore, the processing capacities of next-generation video surveillance systems and the increasing number of closed-circuit television cameras installed in public places are raising concerns about individual's privacy in public spaces too.

In the near future it is expected that cameras will surround us in both public and private spaces. Intelligent monitoring systems threaten individual's right to privacy because of automatic monitoring (Adams & Ferryman, 2013). These systems can retain a variety of information about people habits, visited places, relationships, and so on (Coudert, 2010). It is known that some systems already use facial recognition technology (Goessl, 2012). This way, these systems may build a profile for each citizen in which the people identity and related sensitive information is revealed. Therefore, this evolution of intelligent monitoring systems could be seen as approaching an Orwellian Big Brother, as people may have the feeling of being constantly monitored.

In the light of the above, it is clear that the protection of the individual's privacy is of special interest in telecare applications as well as in video surveillance, regardless whether they operate in private or public spaces. Therefore, privacy requirements must be considered in intelligent monitoring systems by design (Langheinrich, 2001; Schaar, 2010). As aforementioned, smart cameras become essential for AAL applications. Given that security and privacy protection have become critical issues for the acceptance of video cameras, a privacy-aware smart camera would make it possible to use video cameras in realms where they have never been used before. If individual's privacy can be guaranteed through the use of this technology, public acceptance would be increased giving the opportunity of installing these cameras in private environments to replace simpler binary sensors or, most importantly, to develop new telecare services (Chen et al., 2012; Morris et al., 2013; Olivieri, Conde, & Sobrino, 2012). This breakthrough could open the door to novel privacy-aware applications for ambient intelligence (AmI) (Augusto, Nakashima, & Aghajan, 2010), and more specifically in AAL systems for ageing in place (O'Brien & Mac Ruairi, 2009), being beneficial to improve the quality of life and to maintain the independence of people in need of long-term care.

1.1. Related studies

The focus of this review is on the protection of visual privacy. There are valuable reviews about AmI and AAL that have already considered privacy in video and have also highlighted its importance for the adoption of video-based AAL applications (Cardinaux et al., 2011; Cook, Augusto, & Jakkula, 2009). But these works scarcely go into detail about how visual privacy protection can be achieved. Other works from the video surveillance field have also analysed this topic but from a different point of view (Cavoukian, 2013, Senior et al., 2003, 2005). The main threats and risks of surveillance technologies like closed-circuit television cameras, number plates recognition, geolocation and drones are discussed in depth. As a consequence, some guidelines to manage privacy are also proposed, but how to protect visual privacy is not considered. In the same line, Senior and Pankanti (2011) unify their previous works and extend the review of visual privacy not only to video surveillance but also to medical images, media spaces and institutional databases. They consider some technologies to protect visual privacy and provide a classification. As far as we know, this is the first attempt to provide such a classification of protection methods for visual privacy but it is not a comprehensive one. In a more recent work (Winkler & Rinner, 2014), security and privacy in visual sensor networks are reviewed. Although they perform a detailed analysis of the security from several points of view (data-centric, node-centric, network-centric and user-centric), they do not provide an in-depth analysis of privacy protection.

In this survey, we focus on giving an answer to the question of how the visual privacy can be protected, and how such a kind of protection is developed by some of the existing privacy-aware intelligent monitoring systems that have been found in the literature. Because of this, a comprehensive classification of visual privacy protection methods is provided as the main contribution of this work. The remainder of this paper is organised as follows: Section 2 gives an intuitive notion of what visual privacy protection is. A comprehensive review of visual privacy protection methods is presented in Section 3. In Section 4, relevant privacy-aware intelligent monitoring systems are introduced. A discussion of important privacy-related aspects is carried out in Section 5. Finally, a summary of the present work as well as future research directions are given in Section 6.

2. Visual privacy protection

Privacy protection consists in preventing that the information that an individual wants to keep private becomes available to the public domain. In the context of images and videos, we refer to it as visual privacy protection. In this paper, the terms visual privacy and privacy will be used indistinctly, except when indicated.

First of all, it is worth to clarify when individual's privacy needs to be protected. When protecting privacy, it can be differentiated between person's identity and sensitive information which has to be kept in private. Video can convey an enormous amount of information that can be qualified as sensitive. Nevertheless, if sensitive information is present in a video but person's identity is not, there is no privacy loss. The same is true whether person's identity is in a video but without any sensitive information. In both cases, privacy is protected because there does not exist any association or mapping between sensitive information and person's identity.

Another important issue related to visual privacy is which is the sensitive information or region of interest to be protected. In many works only the face is obscured but that is not enough to protect visual privacy. Even when the person's face is obscured, other elements could exist in the image through which person identification may be performed, for instance, using inference channels and previous knowledge (Saini, Atrey, Mehrotra, & Kankanhalli, 2014). Visual cues like clothes, height, gait, and the like can be used to identify the person. For instance, in a pair-wise constraints identification (Chang, Yan, Chen, & Yang, 2006; Chen, Chang, Yan, & Yang, 2009) where faces had been masked, observers were able to identify whether a person in one image was the same one than in a different image. In that test, recognition hit rate was higher than 80%. By using this information and only detecting an image where a privacy breach exists, the person may be identified and tracked in images where privacy was presumably preserved. These visual clues must be considered in order to protect privacy as they affect to the election of which regions of interest have to be protected. So, there is not actually only one region of interest but multiple. A region of interest should be extended to a wider area in some cases, while two or more regions of interest should be created in others.

3. Protection methods

There are different ways to protect and preserve the privacy of individuals appearing in videos and images (see Table 1). Two approaches can be found if we consider the temporal aspect of when a protection method is used, *i.e.* before the image is acquired, or after it. On the one hand, it is possible to prevent others of successfully capturing images in which an individual appears. On the other hand, once an image exists sensitive or private information (*e.g.* faces, number plates, and so on) can be removed (Fig. 1).

According to how privacy is protected, protection methods can be classified in five large categories: *intervention* (Section 3.1), *blind vision* (Section 3.2), *secure processing* (Section 3.3), *redaction* (Section 3.4) and *data hiding* (Section 3.5). Although the latter is often used along redaction methods, it has been added as it is a large field of research itself. In this section, a review of commonly used protection methods to protect visual privacy is presented.

3.1. Intervention

Intervention methods deal with the problem of preventing someone to capture private visual data from the environment. They aim to create capture-resistant spaces. These methods physically intervene camera devices to prevent the acquisition of an image by means of a specialised device that interferes with the camera optical lens. For instance, Patel, Summet, and Truong (2009) developed the BlindSpot system. It locates any number of retro-reflective CCD or CMOS camera lenses around a protected area and, then, it directs a pulsing light at the detected lens, spoiling any images that cameras may record as Fig. 2 shows. Similarly, Harvey (2010) proposed an anti-paparazzi device. It uses an array of high-power LEDs to produce a stream of light at over 12000 lumen. Mitskog and Ralston (2012) have patented a camera blocker for a device with an integrated camera that uses a thin film organic polymer to neutralise camera lens. This camera blocker is reusable and adhesively sticks to any surface without leaving a sticky residue.

Aforementioned approaches are suitable when, once recorded, no control can be enforced over the use of images. However, when enforcement is possible, software-based methods can be used. For instance, the firmware of a camera or an application installed on a smartphone could be responsible of preventing the capture of certain environments, such as artworks in a museum, when the device comes into the range of a bluetooth transmitter (Wagstaff, 2004). Nevertheless, software-based intervention has some drawbacks. First of all, it can be easily thwarted by using a camera without any privacy software installed. Furthermore, users consent and their collaboration is required in order to work successfully. Because systems like these depend on third parties, they are likely doomed to failure. A new privacy legislation that enforces cameras to be in accordance with privacy protocols is needed. Anyway, given that no system is completely secure by nature, even under this assumption, it might be hacked.

3.2. Blind vision

Blind vision has to do with image or video processing in an anonymous way. It addresses privacy related processing issues by means of secure multi-party computation (SMC) techniques that are applied to vision algorithms (Avidan & Butman, 2006). SMC is a subfield of cryptography that enables multiple parties to jointly compute a function in such a way that their inputs and the function itself are not revealed. When applied to vision this means that one party, Bob, could use a vision algorithm of another one, Alice, without enabling Bob to learn anything about Alice's algorithm and, in the same extent, without enabling Alice to learn anything about Bob's images. These methods are useful in order to use third-party algorithms to process data in a privacy-respectful way. For instance, using blind vision techniques, common tasks such as face detection, image matching, object tracking or image segmentation could be done in an anonymous way.

Concerning this, Avidan and Butman (2006) developed a secure classifier that was used in a blind face detector. This classifier is based on an oblivious transfer method and a secure dot-product protocol in order to compute the face detection algorithm. However, it is very time consuming due to the high computational load of the algorithm. Nevertheless, a proposal to accelerate the algorithm by using histograms of oriented gradients is considered at the risk of revealing some information about the original image.

Similarly, Erkin et al. (2009) proposed an efficient algorithm that allows to jointly run the Eigenfaces (Turk & Pentland, 1991) recognition algorithm. This algorithm operates over homomorphically encrypted data. It allows matching an encrypted facial image with a database of facial templates in such a way that the biometric

Table 1
Overview of reviewed papers according to the used visual privacy protection method.

Category	Count	References
Intervention	4	Wagstaff (2004), Patel et al. (2009), Harvey (2010) and Mitskog and Ralston (2012)
Blind vision	5	Avidan and Butman (2006), Erkin et al. (2009), Avidan et al. (2009), Sadeghi et al. (2010) and Shashanka (2010)
Secure processing	8	Erturk (2007), Shashank et al. (2008), Park and Kautz (2008), Ito and Kiya (2009), Upmanyu et al. (2009), Ng et al. (2010), Chaaraoui et al. (2012a) and Zhang et al. (2012)
Redaction: Image filter	10	Hudson and Smith (1996), Zhao and Stasko (1998), Boyle et al. (2000), Neustaedter and Greenberg (2003), Kitahara et al. (2004), Martínez-Ponte et al. (2005), Neustaedter et al. (2006), Zhang et al. (2006), Frome et al. (2009) and Agrawal and Narayanan (2011)
Redaction: Encryption	27	Spanos and Maples (1995), Macq and Quisquater (1995), Agi and Gong (1996), Tang (1996), Qiao et al. (1997), Zeng and Lei (2003), Yang et al. (2004), Boult (2005), Yabuta et al. (2005), Yabuta et al. (2006), Dufaux and Ebrahimi (2006), Dufaux et al. (2006), Chattopadhyay and Boult (2007), Baaziz et al. (2007), Raju et al. (2008), Dufaux and Ebrahimi (2008a, 2008b), Xiangdong et al. (2008), Carrillo et al. (2010), Dufaux and Ebrahimi (2010), Tong et al. (2010), Tong et al. (2011), Dufaux (2011), Sohn et al. (2011), Pande and Zambreno (2013), Li et al. (2013) and Ra et al. (2013)
Redaction: K-same family	7	Newton et al. (2005), Gross et al. (2006), Gross and Sweeney et al. (2006), Bitouk et al. (2008), Gross et al. (2009), De la Hunty et al. (2010) and Lin et al. (2012)
Redaction: Object/people removal	22	Kokaram et al. (1995), Igehy and Pereira (1997), Masnou and Morel (1998), Morse and Schwartzwald (1998), Efros and Leung (1999), Bertalmio et al. (2006), Efros and Freeman (2001), Criminisi et al. (2003), Criminisi et al. (2004), Zhang and Xiao et al. (2005), Chan et al. (2006), Cheung et al. (2006), Shiratori et al. (2006), Wexler et al. (2007), Patwardhan et al. (2007), Whyte et al. (2009), Vijay Venkatesh et al. (2009), Koochari and Soryani (2010), He et al. (2011), Ma and Ma (2011), Ghanbari and Soryani (2011) and Abraham et al. (2012)
Redaction: Visual abstraction	14	Hodgins et al. (1998), Lyons et al. (1998), Tansuriyavong and Hanaki (2001), Fan et al. (2005), Williams et al. (2006), Chen et al. (2007), Baran and Popović (2007), Hogue et al. (2007), Chinomi et al. (2008), Chen et al. (2009), Qureshi (2009), Sadimon et al. (2010), Borosán et al. (2012) and Chen et al. (2014)
Data hiding	12	Petitcolas et al. (1999), Wu (2001), Cox et al. (2002), Yabuta et al. (2005), Zhang et al. (2005), Ni et al. (2006), Yu and Babaguchi (2007), Paruchuri and Cheung (2008), Cheung et al. (2008), Cheung et al. (2009), Paruchuri et al. (2009) and Guangzhen et al. (2010)

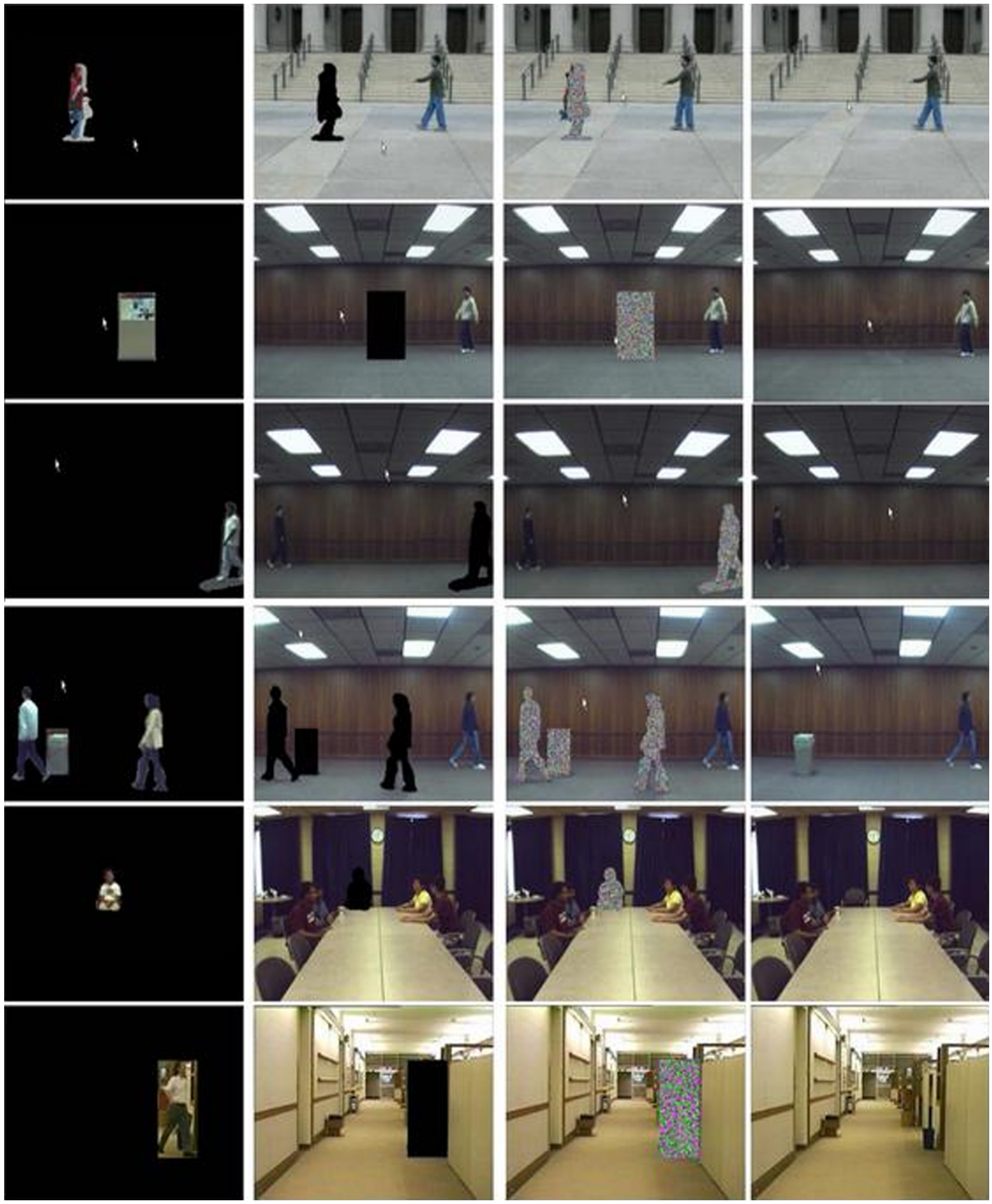


Fig. 1. Some privacy protected sequences of images: The first column shows the sensitive information or region of interest; the second column shows the region of interest replaced by the silhouette; the third column shows the sensitive information scrambled; and the last column shows the sensitive areas inpainted. Reprinted from Paruchuri et al. (2009).

itself and the detection result are kept in secret from the server that performs the matching. Sadeghi, Schneider, and Wehrenberg (2010) improved the communication and computation efficiency of the previous algorithm.

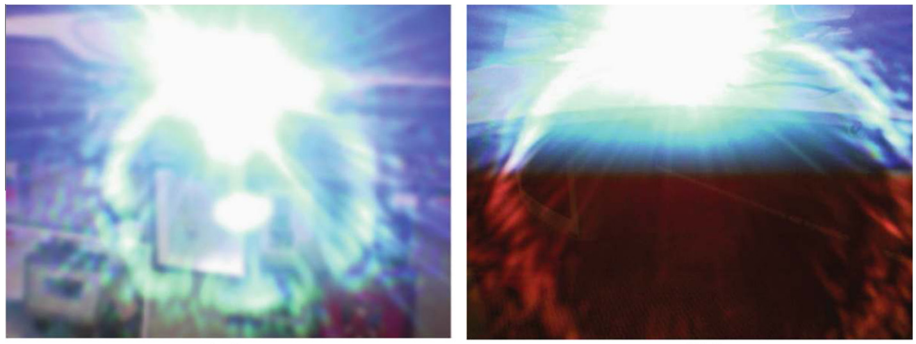
Blind vision algorithms are also used for image matching. Avidan, Elbaz, Malkin, and Moriarty (2009) presented a protocol for image matching related algorithms. Image matching is described as a generalisation of detection and recognition tasks. The described algorithm is built upon a secure fuzzy matching of SIFT attributes, which are treated as 16 character text strings, and it is used in their bag-of-features approach.

A framework for privacy-preserving Gaussian Mixture Model (GMM) computations is proposed by Shashanka (2010). GMMs are commonly used in machine learning for clustering and classification. In computer vision they are mainly used in background subtraction (Zivkovic, 2004) for motion analysis (Chan & Liu, 2009).

3.3. Secure processing

There are other methods that are not based on SMC but that process visual information in a privacy respectful way. They are referred as secure processing in this work. For example, an image matching algorithm for private content-based image retrieval (PCBIR) is proposed by Shashank, Kowshik, Srinathan, and Jawahar (2008). PCBIR is related with similarity search. Conversely to blind vision, privacy is required in one direction because the whole database is usually public, while the query is kept private.

Another possibility is to work with images in a domain that does not reveal visual information. Ito and Kiya (2009) presented an



(a) (b)

Fig. 2. Light beam neutralising a camera lens. A light beam of single colour is used on the left picture, whereas a light beam generated using colour patterns is used on the right picture. Reprinted from Patel et al. (2009). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

image matching algorithm using phase-only correlation (POC) in the frequency domain. This algorithm preserves the visual privacy of the images in a template database. In order to achieve this, all the images of the template database are converted to the frequency domain. Then, a phase scrambling using a one-time key is applied to the discrete Fourier transformation (DFT) coefficients. Afterwards, in order to match a query image with an image of the templates database, the query image is converted to the frequency domain and the matching is done with POC using the DFT coefficients as inputs.

Algorithms that preserve privacy in an implicit manner, for instance, rejecting visual information that is not necessary for the algorithm to work are also considered under the umbrella of secure processing. Ng, minn Ang, and Seng (2010) proposed a privacy-preserving stereoscopic vision algorithm. It preserves privacy by calculating the disparity map using One-Bit Transform (Erturk, 2007). This way, each pixel in the input images is reduced to one bit. In this process, a huge amount of information (colour, texture and so on) is removed in the output images, complicating thus the identification of persons appearing on the images.

A prototype of a privacy-preserving system for recognition of activities of daily living is proposed by Park and Kautz (2008). This prototype relies on the silhouette of detected foreground objects and the motion-map of the frame in order to analyse the activities. This way, if the silhouette and motion-map generation is performed within the camera and it is not possible to access the RGB signal, the visual privacy of the persons inhabiting the environment would be ensured from the algorithm point of view. Similarly, Chaaraoui, Climent-Pérez, and Flórez-Revuelta (2012a) use silhouettes in their efficient approach for multi-view human action recognition based on bag of key poses. By using silhouettes, identifiable information is removed, hence it can also be considered a privacy-aware method. Depth information from RGB-D cameras can be used as a way to preserve privacy (Zhang, Tian, & Capezuti, 2012) as well. Depth data can be obtained from low-cost structured-light cameras like Microsoft Kinect or Asus Xtion, or time-of-flight cameras like Microsoft Kinect 2. Given that no colour information is involved, a depth map visualisation does not enable face recognition and prevents direct extraction of visual clues for person identification.

Upmanyu, Namboodiri, Srinathan, and Jawahar (2009) presented an interesting approach where an efficient framework to carry out privacy-preserving surveillance is proposed. This solution is inspired in a secret sharing scheme adapted to image data. In this framework an image is split into a set of random images in such a way that the shares do not contain any meaningful information about the original image. Despite of shares being distributed between a certain number of servers, computer vision algorithms can be applied securely and efficiently.

3.4. Redaction

Image and video modification or redaction methods are the most common visual privacy protection methods. They modify the sensitive regions of an image such as faces, bodies, number plates, etc. to conceal private information concerning the subjects appearing on it. In order to determine the privacy sensitive regions in which a redaction method operates, computer vision algorithms are used. However, this section focuses only on the application of privacy-preserving methods, therefore it is assumed that sensitive regions are correctly detected.

According to the way in which an image is modified, redaction methods can be classified in several categories. To begin with, there are *ad hoc distortion/suppression* methods (Section 3.4.1). These methods modify the region of interest of an image, either completely removing sensitive information from the image, or modifying the information using common image filters like *blurring* or *pixelating*. By using these filters, sensitive regions are modified in order to make them unrecognisable.

More robust methods like *image encryption* (Section 3.4.2) are also used to conceal the region of interest. Using image encryption techniques the privacy sensitive region of an image is ciphered by a key. Generally encryption techniques based on scrambling (permutation) were commonly used in analogue video, but they can also be used in digital video. Besides of that, it must be taken into account that in the literature scrambling is used as a synonym of encryption in some cases. In this paper, we consider scrambling techniques as a subfield of image encryption.

Another approach to image redaction is *face de-identification* (Section 3.4.3). These methods are focused on de-identifying the faces appearing in an image. Although the aforementioned methods can be used for face de-identification, we focus here on those based on the *k-same* family of algorithms that implements the *k-anonymity* protection model (Sweeney, 2002). These algorithms alter the face of a person in such a way that identity cannot be recognised but facial expressions are preserved.

Finally, some approaches like *object removal* (Section 3.4.4) use inpainting-based algorithms to completely remove sensitive regions of an image by filling the left gap with the corresponding background. Once the image has been inpainted, a visual abstraction of the removed sensitive information can be rendered, like a stick figure, a point, a silhouette, and the like (Chinomi, Nitta, Ito, & Babaguchi, 2008).

Regarding image modification, some questions have to be outlined. Redaction methods cannot modify an image in whatever way. As reported by Hudson and Smith (1996), there is a trade-off between providing privacy and intelligibility in images. For instance, when an image is modified, information needed for image understanding may be also removed. So, a modified image could lack of utility and balancing privacy and intelligibility is needed. In other words, privacy protected images have to retain useful information needed by applications built upon this information, such as telecare monitoring systems or video surveillance. It is also worth mentioning that whereas some of the redaction methods are irreversible, *i.e.* the modification cannot be undone, there are others methods that are reversible so the original version can be recovered if needed. Nevertheless, reversible methods can be developed by using irreversible redaction methods along with data hiding algorithms.

An overview of redaction methods has been given so far. In next subsections, each one of the described categories will be dealt with in depth, focusing on the most representative works and summarising how the presented methods are used.

3.4.1. Image filtering

Redaction methods based on image filtering use common image filters to apply various effects on images in order to modify privacy sensitive regions (Fig. 3). Depending on the application, image filters can be used for obscuring human faces, human bodies, number plates or even background in video conferences (Frome et al., 2009, Kitahara, Kogure, & Hagita, 2004; Martínez-Ponte, Desurmont, Meessen, & Delaigle, 2005; Zhang, Rui, & wei He, 2006). Among the most commonly used filters, blurring (Frome et al., 2009; Neustaedter & Greenberg, 2003; Neustaedter, Greenberg, & Boyle, 2006; Zhang et al., 2006) and pixelating (Boyle, Edwards, & Greenberg, 2000; Kitahara et al., 2004; Neustaedter et al., 2006) stand out.

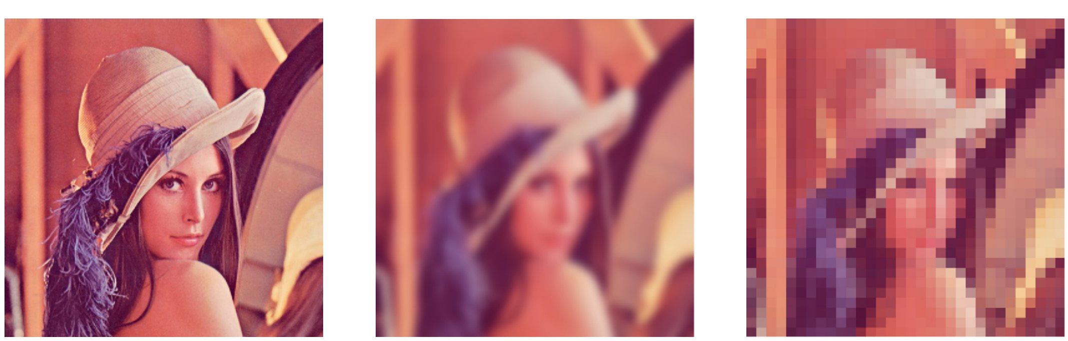
A blurring filter applies a Gaussian function over an image. This function modifies each pixel of an image using neighbouring pixels. As a result, a blurred image is obtained in which the details of sensitive regions have been removed. Blurring is widely used in Google Street View (Frome et al., 2009) to modify human faces and number plates. A pixelating filter divides an image into a grid of eight-pixel wide by eight-pixel high blocks. The average colour of the pixels of each block is computed and the resultant colour is assigned to all of the pixels belonging to that block. As a result, an image where the resolution of sensitive regions have been reduced is obtained. Pixelating is commonly used in television to preserve the anonymity of suspects, witnesses or bystanders. However, it is vulnerable to some kind of attacks such as integrating pixels along trajectories over time that may allow partly recovering of the obscured information.

Lander, Bruce, and Hill (2001) evaluated the effectiveness of pixelating and blurring in videos and static images. Results indicated that participants were still able to recognise some of the faces. Similarly, Newton, Sweeney, and Malin (2005) showed that pixelating and blurring alike filters do not thwart recognition software. Training a parrot recogniser using the same distortion as the probe on gallery images, high recognition rates are obtained (near 100%) despite looking somewhat de-identified to humans.

3.4.2. Encryption of videos and images

Image and video encryption methods encode imagery data in such a way that the original data becomes unintelligible as can be observed in Fig. 4. The main goal of these methods is reliable security in storage and secure transmission of content over the network. Usually, security in cryptographic algorithms resides in the strength of the used key instead of in keeping the algorithm private. By using encryption, a distorted video that unauthorised viewers cannot visualise is obtained. Only users who have the proper key for decryption can visualise it. Through the inverse operation and the used key, the ciphered data can be deciphered in order to retrieve the original images. This is named conditional access through encryption. Encryption methods operate over the whole frame or a delimited region of all the video frames. Although such methods do not provide a balance between privacy and intelligibility, they enable to perform data analysis over unprotected data once authorisation and required permissions have been granted. In such cases, privacy would be protected until access to raw data is eventually requested.

Generally naïve video encryption algorithms have treated the compressed video bitstream as text data, therefore encrypting the entire video bitstream. Hence, commonly used encryption algorithms such as Data Encryption Standard (DES), Rivest's Cipher (RC5), Advanced Encryption Standard (AES), Rivest, Shamir and Adleman (RSA) and so on, have been used. These algorithms guarantee the highest security level but, unfortunately, they are not suitable for real-time video encryption because they are very time consuming (Yang, Bourbakis, & Li, 2004; Pande & Zambreno, 2013). Due to this, selective encryption algorithms have been proposed (Spanos & Maples, 1995). These algorithms keep using text-based encryption but encrypt only a selected part of the video bitstream so as to get real-time encryption. Other encryption algorithms have also been proposed for real-time encryption, namely light-weight encryption algorithms (Zeng & Lei, 2003). These algorithms are suitable for real-time applications because, when encrypting, they use a simple XOR cipher or only encrypt some bits of the video bitstream. Thereby, they are much faster than the first ones. Finally, there are methods based on scrambling (Tang, 1996). Traditional scrambling methods modify an analogue video signal like those found on closed-circuit television cameras to make it unintelligible. However, with the proliferation of digital video cameras, scrambling techniques are also applied to digital videos in the field of video encryption. Mainly, scrambling algorithms are based on



(a) Lena (b) Lena with Gaussian Blur (c) Lena Pixelation

Fig. 3. Several examples of image filtering methods: (a) the original image without any modification, (b) image modified by applying a Gaussian Blur filter, and (c) image modified by applying a pixelating filter.

Fig. 4. Two examples of an encrypted image where the face of the person is considered the sensitive region. Reprinted from Boult (2005).

permutation only methods in which transformed coefficients are then permuted in order to distort the resulting image.

Each one of these methods can operate in a specific domain, like the spatial domain, frequency domain (transform domain) or code-stream domain (compressed video). Furthermore, it is important to note that light-weight encryption and scrambling-based methods are less secure than naïve encryption. For instance, scrambling video in the spatial domain is subject to efficient attacks (Macq & Quisquater, 1995). Generally these algorithms trade-off security for encryption speed. However, compared to blurring and pixelating, scrambling approaches as in Dufaux and Ebrahimi (2008a) are successful at hiding identity (Dufaux & Ebrahimi, 2010; Dufaux, 2011).

Regarding when encryption is performed, there are several approaches: prior to encoding, after encoding or during encoding. Each approach has advantages and disadvantages. Prior-to-encoding encryption is a very simple method that works with the original image independently from the used encoding scheme. However, it significantly changes the statistics property of the video signal, resulting in a less efficient compression later. Regarding after-encoding encryption, the compressed code-stream is encrypted after video encoding. The resulting encrypted and compressed code-stream could hardly be reproducible in a standard player, and it could even cause the player to crash. However, it avoids to fully decode and re-encode the video. Finally, during-encoding encryption has the advantage of a fine-grained control over the encoding process but it is closely linked to the used video encoding scheme.

Next, some of the selective and light-weight encryption as well as scrambling methods found in the literature will be analysed.

Concerning selective encryption, AEGIS (Spanos & Maples, 1995) is an algorithm that uses DES to encrypt only the I-frames of the MPEG video bitstream. By ciphering I-frames, the needed B-frames and P-frames cannot be reconstructed either. However, the algorithm is not completely secure due to partial information leakage from the I-blocks in P and B frames (Agi & Gong, 1996). Similarly, the video encryption algorithm proposed by Qiao et al. (1997) works with I-frames. It divides them in two halves that are XORed and stored in one half. One of the half is encrypted using DES algorithm. Although this algorithm is secure, it is not suitable for real-time applications. Raju, Umadevi, Srinathan, and Jawahar (2008) analyse the distribution of the DCT coefficients (DC and AC) of compressed MPEG videos in order to develop a computationally efficient and secure encryption scheme for real-time applications. DC and AC coefficients are managed differently regarding their visual influence, and electronic code block and cipher block chaining modes are interleaved to adapt the encryption process to the video data. The described scheme uses RC5 for encrypting DCT coefficients. Boult (2005) used DES and AES to encrypt faces in JPEG images during compression in their privacy approach through invertible cryptographic obfuscation. The information required for the decrypting process is stored inside the JPEG file header. Although this information can be publicly read, it cannot be used without the private key. Chattopadhyay and Boult (2007) used this technique for real-time encryption, using uCLinux on the Blackfin DSP architecture. Similarly, an encryption scheme for JPEG images is used by Ra, Govindan, and Ortega (2013). The JPEG image is divided into two parts, one public and one private. The first one is unaltered, whereas in the second one the most significant DC coefficients are encrypted during the encoding process after the quantisation step. This approach is designed to obtain JPEG-compliant images that can be sent to photo sharing services under storage and bandwidth constraints.

As for light-weight video encryption, Zeng and Lei (2003) presented an efficient algorithm for H263 that operates in the frequency domain. Bit scrambling is used to transform coefficients and motion vectors during video encoding without affecting the compression efficiency. By using this method each frame of the resulting video is completely distorted. A cryptographic key is used to control the scrambling process, thereby authorised users will be able to undo the scrambling using the key. A similar video encryption algorithm for MPEG-4 is proposed by Dufaux and Ebrahimi (2006) where security is provided by pseudo-randomly inverting the sign of selected transform coefficients (frequency domain) corresponding to the regions of interest. The encryption process depends on a private key that is RSA encrypted and inserted in the stream as metadata. In this method the amount of distortion introduced can be adjusted from merely fuzzy to completely noisy. This method is deeply explained for the case of Motion JPEG 2000 in Dufaux et al. (2006), and for the case of H264/AVC in Dufaux and Ebrahimi (2008a). Concerning the latter, Tong, Dai, Zhang, and Li (2010) made a proposal to correct the drift error produced in H264/AVC during video encryption. The described method is also used by Baaziz, Lolo, Padilla, and Petngang (2007) in an automated video surveillance system. Dufaux and Ebrahimi (2008b) presented an extension of their previous work based on code-stream domain scrambling. The scrambling is performed after the MPEG-4 encoding process directly on the resulting code-stream output by the camera. This way, it avoids to decode and re-encode the video, saving computational complexity. An encryption scheme for JPEG XR (Srinivasan, Tu, Regunathan, & Sullivan, 2007) working in the frequency domain is proposed by Sohn, De Neve, and Ro (2011). It uses subband-adaptive scrambling for protecting face regions. Concretely, different scrambling techniques are used for each subband: random level shift for DC subbands, random permutation for LP subbands, and random sign inversion for HP subbands. A different encryption scheme based on compressive sensing (CS) (Candes, Romberg, & Tao, 2006; Donoho, 2006) and chaos theory is proposed by Tong, Dai, Zhang, Li, and Zhang (2011). This method scrambles sensitive regions of a video by using block-based CS sampling on their transform coefficients during encoding. The scrambling process is controlled by a chaotic sequence used to form the CS measurement matrix. In order to prevent drift error they also use their previous method.

Finally, regarding scrambling, Tang (1996) proposed an encryption algorithm that works in the frequency domain and uses the permutation of the DCT coefficients in order to replace the zig-zag order. Specifically, instead of using the zig-zag order to map the 8x8 block to a 1x64 vector, a random permutation list is used. A different approach is proposed by Yabuta, Kitazawa, and Tanaka (2005) where scrambling is performed in the space domain before encoding. The scrambling only affects the moving regions and is performed by randomly permuting pixels. Before being scrambled, the original moving regions are encrypted with AES, and embedded inside the masked image using digital watermarking (Petitcolas, Anderson, & Kuhn, 1999). Thereby, the scrambled regions can be recovered later if needed. An improved version of this method is presented in Yabuta, Kitazawa, and Tanaka (2006) being able of real-time encoding, and decoding only one specific object by exploiting object tracking information. Xiangdong, Junxing, Jinhai, and Xiqin (2008) presented a novel encryption scheme that relies on chaos theory. It works in the space domain and prior to video encoding. It permutes the pixels of an image row by row using a chaotic sequence of sorted real numbers as a key. This key can be used to de-scramble the image when necessary. A similar approach that sorts the chaotic sequence as Vigenère cipher is proposed by Li, Zhao, Qu, and Wang (2013). Carrillo, Kalva, and Magliveras (2010) proposed an encryption algorithm working in the spatial domain. Before encoding, pseudo-random permutations are applied to the pixels. A secret pass phrase which controls the permutation process is used as key. This algorithm is independent of the used compression algorithm and robust to transcoding at the cost of an increase in bitrate depending on the percentage of encrypted blocks.

3.4.3. Face de-identification

Face de-identification consists in the alteration of faces to conceal person identities. The goal is to alter a face region in such a way that it cannot be recognised using face recognition software. In order to achieve this, the faces appearing in images are modified by using some of the aforementioned methods such as image filtering previously discussed in Section 3.4.1. Nevertheless, there are cases in which privacy must be preserved while images must still keep their capacity of being analysed. In such cases, it is then necessary to balance privacy and intelligibility. Concerning this, there are methods in the literature that consider this trade-off, and some of them will be reviewed in this section.

The k-Same family of algorithms (Fig. 5) is one of the most recent and commonly used algorithms for face de-identification. K-Same was first introduced by Newton et al. (2005). Intuitively, k-Same performs de-identification by computing the average of k face images in a face set. Then, all of the images of the given cluster are replaced by the obtained average face. Using this algorithm, a de-identified face is representative of k members of the original used face set. This way, if the probability of a de-identified face of being correctly recognised by a face recognition software is no more than $\frac{1}{k}$, it is said that this algorithm provides *k-anonymity* privacy protection (Sweeney, 2002). However, despite the fact that this formal model can preserve privacy, there are no guarantees of the utility of the data.

Gross, Sweeney, de la Torre, and Baker (2006) proposed some extensions to the k-Same. On the one hand, an algorithm named k-Same-Select that extends the k-Same algorithm is presented in Gross, Airoldi, Malin, and Sweeney (2006). It guarantees the utility of the data, for example, by preserving facial expressions or gender in face images. This algorithm divides the face image set into mutually exclusive subsets using a data utility function and, then, it uses the k-Same algorithm on the different subsets. On the other hand, the k-Same-M algorithm is introduced in Gross and Sweeney et al. (2006) to fix a shortcoming of the k-Same-Select algorithm. Appearance-based algorithms, like k-Same-Select, work directly in the pixel level producing sometimes alignment mismatch that leads to undesirable artefacts in the de-identified face images. In order to overcome this issue k-Same-M relies on active appearance model (Cootes, Edwards, & Taylor, 1998; Cootes, Edwards, & Taylor, 2001), a statistical and photo-realistic model of the shape and texture of faces. This model is also used by De la Hunty, Asthana, and Goecke (2010) for real-time facial expression transferring from one's person face to another. Despite this method is not used for face de-identification, it could be useful to transfer expressions from an already de-identified face to another one.

The k-Same-based algorithms have some constraints. For instance, if a subject is represented more than once in the dataset then k-Same does not provide k-anonymity privacy protection. In order to address this shortcoming, Gross, Sweeney, Cohn, Torre, and Baker (2009) proposed a multi-factor model which unifies linear, bilinear and quadratic models. By using a generative multi-factor model, a face image is factorised into identity and non-identity factors. Afterwards, the de-identification algorithm is applied and the de-identified face is reconstructed using the multi-factor model.

A different approach is described by Bitouk, Kumar, Dhillon, Belhumeur, and Nayar (2008), where faces are automatically replaced in photographs. A large library of 2D faces downloaded from the Internet is built according to appearance and pose. In order to replace a detected face in an image, a similar candidate to the input is selected from the library. Lin, Wang, Lin, and Tang (2012) presented a similar work for face swapping where a personalised 3D head model from a frontal face is built, thereby any pose can be rendered by directly matching with the face that wants to be substituted.

3.4.4. Object/people removal

Object and people removal deals with concealing persons or objects appearing in an image or a video in such a way that there are not trails of them in the resulting modified version (Fig. 6). For the sake of clarity, the term 'object' will be used to refer both person and object indistinctly. When removing an object from an image a gap is left. This gap has to be filled in order to create a seamless image. Inpainting methods are used to repair regions with errors or damages. Inpainting consists in reconstructing missing parts in such a way that the modification is undetectable. Information from surrounding area is used to fill in the missing areas. Therefore, these methods can be used for visual privacy to remove people that do not commit suspicious activities from video surveillance recordings, removing inactive participants from the video stream of a video conference, concealing people from images in online social networks, and so many other applications. However, due to computational restrictions, inpainting methods are rarely used in real-time applications running on commodity hardware (Granados et al., 2012).

Inpainting can be divided into two large groups: image inpainting (Bertalmio, Sapiro, Caselles, & Ballester, 2006) and video inpainting (Abraham, Prabhavathy, & Shree, 2012; Kokaram, Morris, Fitzgerald, & Rayner, 1995). This distinction is mainly due to the content nature. While in an image is only needed to ensure spatial consistencies, in video, temporal consistencies between all of the frames have to be ensured as well.

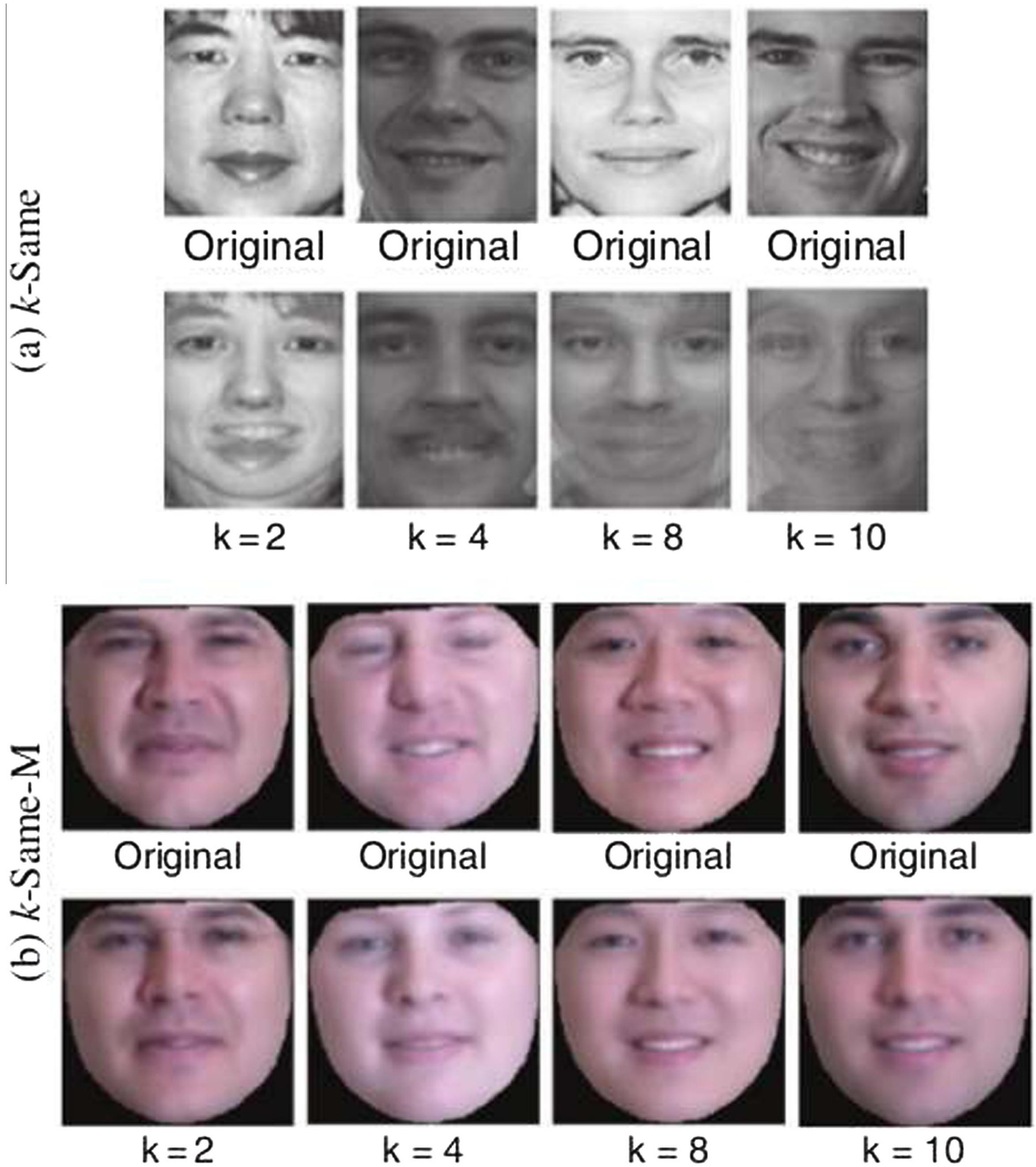


Fig. 5. Several examples of de-identified face images: (a) Faces de-identified by using the k-Same algorithm where some ghosting artefacts appear due to misalignments in the face set. (b) Faces de-identified by using k-Same-M algorithm. Reprinted from Gross et al. (2009).

Regarding image inpainting, there are several approaches: texture synthesis (Efros & Freeman, 2001; Efros & Leung, 1999; Igehy & Pereira, 1997), partial differential equations (PDE) inspired algorithms (Bertalmio et al., 2006; Chan, Shen, & Zhou, 2006; Masnou & Morel, 1998; Morse & Schwartzwald, 1998), and exemplar based (Criminisi, Perez, & Toyama, 2003, 2004; He, Bleyer, & Gelautz, 2011; Koochari & Soryani, 2010; Ma & Ma, 2011; Whyte, Sivic, & Zisserman, 2009). In texture synthesis, a synthetic texture derived from one portion of the image is used to fix another portion. This synthetic texture is a plausible patch that does not have visible seams nor repetitive features. Algorithms based on texture synthesis are able to fill in large regions but at the cost of not preserving linear structures. PDE inspired algorithms reconstruct the gap using geometry information to interpolate the missing parts. A diffusion process propagates linear structures of equal grey value (isophotes) of the surrounding area into the gap region. Although these algorithms preserve well linear structures, the diffusion process introduces some blurring when filling in large regions. Finally, exemplar-based methods are of particular interest. Instead of generating synthetic textures, these methods operate under the assumption that the information that is necessary to complete the gap has to be fetched from nearby regions of the same image. They generate new textures by searching for similar patches in the image with which the gap is filled. Moreover, some exemplar-based algorithms also search the needed information in databases of millions of images in order to complete the remaining information (Whyte et al., 2009). Furthermore, these algorithms often combine the advances of texture synthesis and isophote-driven inpainting by a priority-based mechanism which determines the region filling order, thereby reducing blurring caused by prior techniques.

(a) Real image

(b) Modified image

Fig. 6. An example of a people removal method where the person has been manually selected in the real image (a), and then automatically removed in the second image (b) by filling the region concerning the person using an exemplar-based image inpainting method. Reprinted from Criminisi et al. (2004).

Regarding video inpainting, some of the first straightforward approaches tried to apply image inpainting methods to individual images of the underlying video data. However, they did not take full advantage of the temporal correlation of video sequences. Due to this, the previous methods are often modified in order to be adapted to sequences of images, reconstructing a given frame by interpolating missing parts from adjacent frames. These methods can be classified into patch-based methods (Ghanbari & Soryani, 2011; Patwardhan, Sapiro, & Bertalmio, 2007; Shiratori, Matsushita, Tang, & Kang, 2006; Wexler, Shechtman, & Irani, 2007) and object-based methods (Cheung, Zhao, & Venkatesh, 2006; Vijay Venkatesh, Cheung, & Zhao, 2009; Zhang, Xiao, & Shah, 2005). As patch-based methods are unable to perform both spatial and temporal aspects simultaneously, object-based methods were introduced in order to overcome these constraints.

3.4.5. Visual abstraction/object replacement

Object replacement involves the substitution of objects (or persons) appearing in an image or video by a visual abstraction (or visual model) that protects the privacy of an individual while enabling activity awareness. As far as we know, the term 'visual abstraction' was early coined by Chinomi et al. (2008) to refer to a visual model that abstracts a replaced object. Common visual models could be a point, a bounding box, a stick figure, a silhouette, a polygonal model, and many others as seen in Fig. 7. Visual abstraction may be obtained in a variety of ways. Hence, there is an overlap with some of the previous reviewed methods, such as image filtering or face de-identification, that could also be used as visual models. Object replacement does not necessarily imply removing the object to be replaced, but quite often the aforementioned techniques in Section 3.4.4 intervene. In such cases, an object removal method is applied, followed by the rendering of the visual model over the inpainted image. The abstract object is often located in the same relative position, pose, and orientation as the original object.

Although it depends on the used visual model, by using a proper abstract representation of an object, the information of the scene remains useful so as visual analysis can still be carried out, for instance, to assess the underlying activity before an accident in a home risk detection service. Then, the activity can be analysed without violating the right to privacy of people appearing in the image because it is not possible to directly identify them. However, as reported by Hodgins, O'Brien, and Tumblin (1998), some works in human motion perception showed that viewers easily recognise friends by their gaits, as well as the gender of unfamiliar persons when using moving light displays (Johansson, 1973). Concerning this, it seems that motion is essential for identifying human figures. Furthermore, apparently the geometric model used for rendering human motion affects the viewer's perception of motion. For example, in experiments carried out by Hodgins et al. (1998), subjects were able to better discriminate motion variations using the polygonal model than they were with the stick figure model. Hence, a study to determine how visual abstraction techniques actually preserve privacy must be done.

Different works employ visual abstraction and object replacement. For example, a silhouette representation can be used as a visual abstraction of a person (Tansuriyavong & Hanaki, 2001). This removes information about textures while maintaining the shape of the person, thereby complicating the identification. A silhouette representation is used, for instance, to preserve individual's privacy in a fall detector and object finder system (Williams et al., 2006). Another representation can be obtained by using an edge motion history image (Chen, Chang, Yan, & Yang, 2007, 2009). By using this pseudo-geometric model, the whole human body is obscured and the person looks like a ghost in the final image. This method detects edges of the body appearance and motion, accumulating them through the time, in order to smooth the noise, and partially preserve body contours. Chinomi et al. (2008) proposed twelve operators for visual abstractions: *as-is*, *see-through*, *monotone*, *blur*, *mosaic*, *edge*, *border*, *silhouette*, *box*, *bar*, *dot* and *transparency*. In order to choose among one of them, the relationship between the subject being monitored and the viewer is taken into account. In addition to the representation that have been seen so far, 3D avatars can be used too. Avatar creation (Sadimon, Sunar, Mohamad, & Haron, 2010) is a very interesting field with straightforward application in visual abstraction as well. Lyons, Plante, Jehan, Inoue, and Akamatsu (1998) presented a method that uses

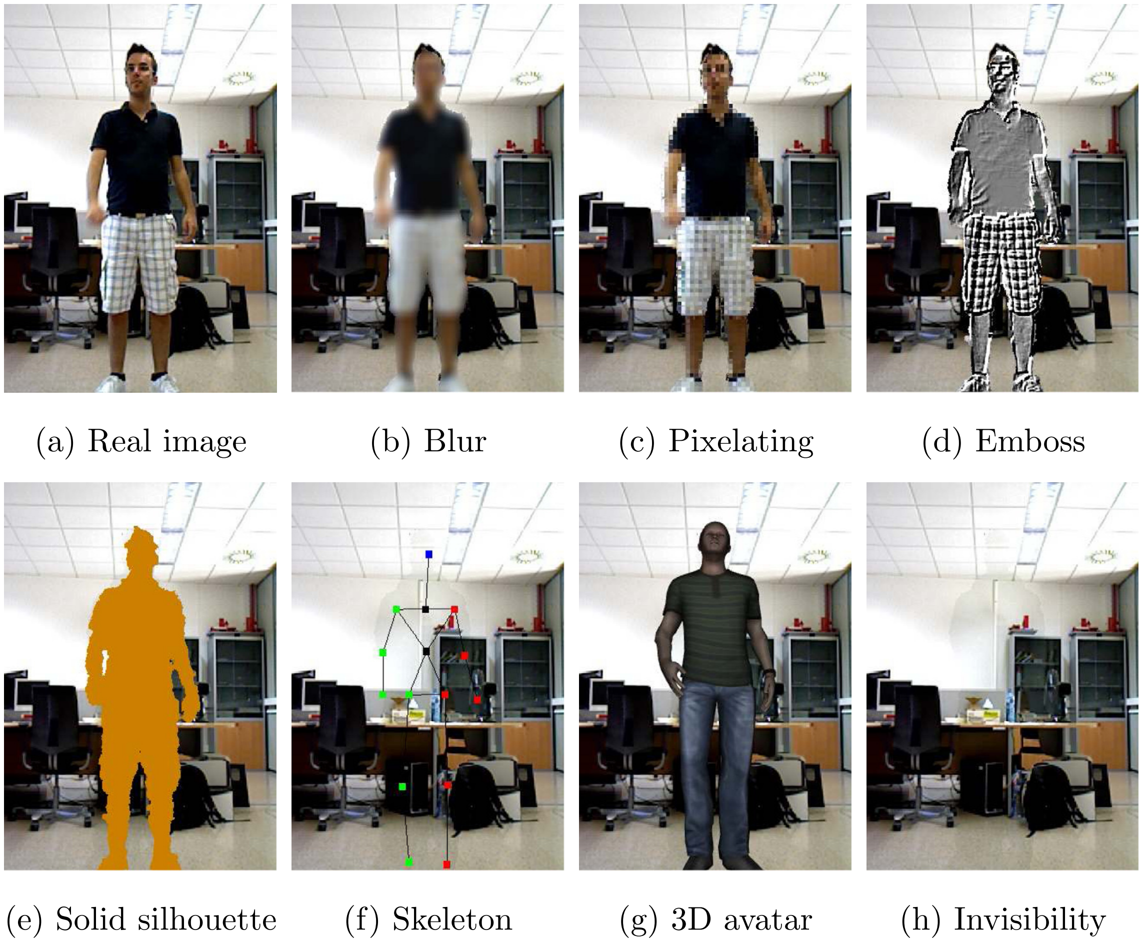


Fig. 7. Several examples of visual abstractions where the person in the real image (a) has been replaced by a visual model.

automatic face recognition and Gabor wavelet transform for creating avatars. It automatically extracts a face from an image, and create a personalised avatar by rendering the face into a generic avatar body. However, given that the avatar maintains recognisable aspects of the face, privacy would not be protected. Hogue, Gill, and Jenkin (2007) proposed a method based on 3D reconstruction for whole body avatar creation. They use a portable stereo video camera to extract the geometry of the person and create a 3D model. A similar and more recent work is proposed by Chen, Dang, Cheng, and Xu (2014), where two low-cost RGB-D cameras are used to scan the whole body of a person in order to create a mesh model. Although the resulting textured model of both methods do not protect privacy, they can constitute a building block for other methods where generic 3D avatars are used to replace persons, as proposed by Fan, Luo, Hacid, and Bertino (2005). Furthermore, automatic rigging (Baran & Popović, 2007; Borosán, Jin, DeCarlo, Gingold, & Nealen, 2012) could be used to animate 3D models, enabling the use of customised 3D models in real-time. Hence, opening the door to new protection methods that transform or deform mesh models. Finally, the use of object-video streams for preserving privacy is proposed by Qureshi (2009). A separated video bitstream is generated for each foreground object appearing in the raw video. Thereby, the original video can be reconstructed from object-video streams without any data loss. During reconstruction, each video bitstream can be rendered in several ways, for instance, obscuring people identities using a silhouette representation or just not showing an object-video stream at all.

3.5. Data hiding based methods

In order to protect privacy, there are redaction methods that apart of modifying the region of interest, they embed the original information inside of the modified version so as to be retrieved in the future if needed (Cheung, Venkatesh, Paruchuri, Zhao, & Nguyen, 2009). These redaction methods make use of data hiding techniques (Petitcolas et al., 1999) to develop reversible methods when the underlying redaction does not support it. In Fig. 8, a classification of data hiding methods is presented. Concerning the terminology used in data hiding, the embedded data is the message that will be sent secretly. It is often hidden in another message referred to as cover message whose content can be text, audio, image or video. As a result of a hidden process, a marked message is obtained.

Generally data hiding techniques are used for steganography, digital watermarking and fingerprinting. Steganography is a practice that consists in concealing a secret message as embedded data inside a cover message. The hiding process is often controlled by a key in order to allow the recovery of the secret message to parties that know it. Regarding digital watermarking and fingerprinting, they both use steganography techniques but are focused on copyright protection. On the one hand, digital watermarking encodes information about the owner of an object by means of a visible pattern (*e.g.* a company logo) or hidden information. On the other hand, fingerprinting is used to embed hidden serial numbers that uniquely identify an object in such a way that the owner of the copyright can detect violations of licence agreements. Data hiding

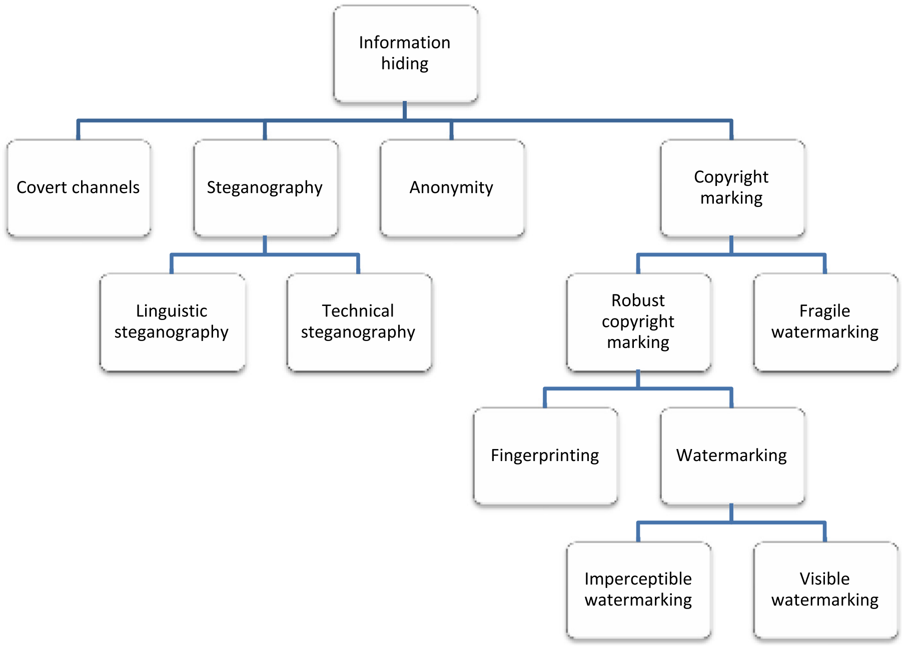


Fig. 8. A classification of data hiding techniques according to Petitcolas et al. (1999).

methods have to provide different kind of features in terms of capacity, perceptibility and robustness (Cox, Honsinger, Miller, & Bloom, 2002). The main difference between steganography and digital watermarking (and fingerprinting) is their application. Whereas the former aims for imperceptibility to human vision, the latter is focused on the robustness.

Invisible watermarking can be used for privacy protection purposes. Instead of embedding information about the owner, the original video is embedded inside of the privacy protected version of it. Due to this, data hiding methods that provide large embedding capacity are required. Furthermore, if the reversibility of the hiding process is considered, irreversible and reversible data hiding techniques can be found (Ni, Shi, Ansari, & Su, 2006). In the former, the cover video cannot be fully restored after the hidden data is extracted, but it usually produces higher hiding capacity. In the latter, the cover video can be fully restored, thereby maintaining the authenticity of the original video. Nevertheless, reversible data hiding is not required in privacy protection because the cover video is discarded once the embedded data has been extracted.

Some of the privacy protection methods that make use of non-reversible and imperceptible watermarking are described next. For instance, Yabuta et al. (2005) extract moving objects of a motion JPEG video. These regions are scrambled in the original video, resulting in a privacy protected video that is used as the cover video. Then, the extracted objects are JPEG compressed and encrypted with AES, followed by a hiding process that embeds them into the least significant bits of middle frequency DCT coefficients of the cover video. The added perturbation is small and does not visually affect the reconstructed image. A similar approach is proposed by Zhang, Cheung, and Chen (2005), where foreground objects are extracted from a H263 video, compressed and encrypted as a regular video bitstream. The gaps left by the foreground objects are inpainted in the original video using background replacement. Then, the encrypted foreground objects are embedded in the DCT coefficients of the inpainted video considering the perceptual quality of the marked video. However, one drawback is the increased bitrate of the resultant video. Yu and Babaguchi (2007) presented a method for hiding a face into a new generated face. It considers the face as the information to be embedded. Initially, an active appearance model is built and faces are characterised as parameters according to this model in order to reduce the payload of the embedded information. Then, the face appearing in the original MPEG2 video is masked. Finally, face parameters are embedded into DCT coefficients of the video following the quantised index modulation scheme (Cox et al., 2002). A novel data hiding algorithm for M-JPEG that minimises both the output perceptual distortion and the output bitrate is proposed by Paruchuri and Cheung (2008). The algorithm identifies the optimal locations to hide data by selecting the DCT coefficients that minimise a cost function that considers both distortion and bitrate. The person appearing in a video (privacy data) is extracted and removed from it. Then, the privacy data is embedded into the selected DCT coefficients of the inpainted video. The proposed algorithm is used by Cheung, Paruchuri, and Nguyen (2008) but modifying the rate-distortion scheme to determine the optimal number of bits to be embedded in each DCT block. This last version is used for a reversible data hiding scheme in a video surveillance prototype (Cheung et al., 2009; Paruchuri, Cheung, & Hail, 2009). Guangzhen, Yoshimichi, Xiaoyi, Naoko, and Noboru (2010) presented a different method for JPEG images. It uses scaling and wavelet coefficients generated by DWT. The former is used to build a low-resolution image where privacy information appears pixelated, whereas the latter is used along scaling coefficients to jointly recover the original image. In this way, wavelet coefficients represent only the private information that needs to be embedded. They are embedded in the DCT coefficients of the low-resolution image after quantisation via amplitude module modulation (Wu, 2001).

4. Privacy-aware intelligent monitoring systems

This section introduces some of the existing expert and intelligent systems in the video surveillance and AAL fields that take privacy into account and, thereby, are developed as a framework to protect it. Some of the reviewed systems focus only on the way in which private date is managed, whereas others use redaction methods to protect individual's privacy before imagery data is stored, distributed, shared or visualised. Because of this, some privacy design questions which such systems should address are drawn (Senior et al., 2005; Yu et al., 2008):

- What data is captured?
- Has the subject given consent?
- How does the subject specify the privacy preferences?
- What form does the data take?
- Who sees the data?
- How long is the data kept?
- How is the data used?
- How raw is the data?
- Who and what should be protected?
- How should privacy be protected in a video surveillance system?
- How can privacy be protected without losing utility?

The answers to these questions lead to privacy policies and technical aspects that such systems should cope with in one way or another. In the same line, it is also important to question when redaction is performed. As stated by Senior (2009), there are mainly several locations where redaction can be performed for a general video-based monitoring architecture made up of cameras, video processor, database and user interfaces. Concretely, these locations include all of the mentioned parts but cameras because of technical issues that may arise in legacy systems. Nevertheless, cameras have been also included here from a conceptual point of view, so there are a total of four locations. In such an architecture, data flows from video cameras to user interfaces, crossing through the video processor and the database. Furthermore, depending on privacy policies of subjects, viewers permission, and others, both redacted and unredacted data may have to be displayed. Next, we enumerate the several locations where redaction can be applied:

1. *User Interface*. This location is the most insecure. Data crosses the system without being protected until it reaches the user interface. Then, redaction is carried out by the user interface and the protected information is presented to the viewer. An advantage of this approach is that redacted data does not require to be stored. However, metadata information needs to be delivered with the raw data.
2. *Database*. When the user interface requests the database for some information to view, the latter can redact it. Thereby, redacted data does not require to be stored but it involves additional processing because the same data may be redacted multiple times. Besides of that, latency issues may arise due to extra processing.
3. *Video Processor*. Redaction can be performed by the video processor before storing. It analyses video bitstreams to detect activities and extract useful information. Nevertheless, bandwidth and storage requirements are increased because both redacted and unredacted data need to be sent to the database.
4. *Video Camera*. A privacy-aware smart camera can redact sequences of images itself. This is the earliest possible stage in which redaction can be applied. Similarly to the video processor, in this location bandwidth and storage requirements are increased too.

Several redaction locations can be combined to enhance security as in *double redaction* also proposed by Senior (2009), in which privacy protection is applied at the earliest stage as well as in other locations. Video is decomposed in multiple information streams containing the private data which are encrypted and flow through the system to the database. The information can be recombined later, when needed by an authorised viewer. By using double redaction more than one level of privacy protection can be provided, where each one could be suitable for different applications. This way, some visual information could not be protected but securely stored, whereas other information could be redacted according to several factors like subject, viewer, ongoing activity and so on.

Next, some of the existing systems found in the literature are described.

CoMedi (Coutaz, Bérard, Carraux, Astier, & Crowley, 1999) is a media space prototype that facilitates remote informal communication and group awareness while assuring privacy protection. A porthole display with fish-eye feature is used in order to provide awareness of the remote activities that are being carried out. It shows the personal information that the corresponding remote user has previously accepted to reveal without losing awareness about peripheral activities. Regarding privacy, this prototype uses a face tracker and a privacy filter based on eigenspace coding in order to filter the captured faces not belonging to the image set of 'socially correct' faces.

NeST (Fidaleo, Nguyen, & Trivedi, 2004), the networked sensor tapestry, is a general architecture to manage the broad range of distributed surveillance applications. It is scalable and is composed of software modules and hardware components. The core component of a NeST server is the privacy buffer. It utilises programmable plug-in privacy filters which operate on data coming from sensors in order to prevent access to it or to remove personal identifiable information. These privacy filters are specified using a privacy grammar that is able of connecting multiple low-level privacy filters to create arbitrary data-dependent privacy definitions. Moreover, the privacy of the monitored subjects is also considered. In this sense, the NeST architecture integrates individuals' privacy (according to behaviours) denying access to specific information to some or all of the modules or operators. NeST features secure sharing, capturing, distributed processing and archiving of surveillance data.

Stealth Vision (Kitahara et al., 2004) is an anonymous video capturing system that protects the privacy of objects by pixelating or blurring their appearance. It determines private areas of captured videos from a mobile camera using 3D space models. The 3D position of the target object is estimated by a homographic transformation using images coming from a static overhead camera. Afterwards, a calibrated mobile camera estimates the privacy area by projecting the 3D models onto the captured image plane. Finally, the privacy area is protected. RFID sensors are employed in order to indicate which persons or objects are allowed to be captured.

PriSurv (Chinomi et al., 2008) is a system that uses visual abstract representations to protect individual's privacy. It is composed of several modules: analyser, profile generator, profile base, access controller, abstractor and video database. In order to identify subjects, RFID tags and image processing is used. This system is focused on small communities where a certain number of members are registered and they monitor each other. Given that the sense of privacy is different for everyone, privacy policies are determined according to the closeness between subjects in the video and viewers monitoring. Hence, the appearance of subjects is opened to close viewers while it is hidden to the viewers that subjects feel distant from. Depending on this closeness, different abstract representations are chosen.

Respectful Cameras (Schiff, Meingast, Mulligan, Sastry, & Goldberg, 2009) is a prototype system focused on addressing privacy concerns. This system works by detecting coloured markers in real time that people wear such as hats or vests in order to indicate their privacy preferences. This are expressed as their will of remaining anonymous. The system obscures the face of a person when a marker has been detected.

CareLog and BufferWare (Hayes & Truong, 2009) are two systems based on the concepts of the selective archiving model.

Under this model, data is constantly buffered but an explicit input is required in order to archive it, otherwise data is lost. It represents a compromise whereby people can negotiate their own policies concerning control, privacy, information access and comfort. CareLog is a system aimed at education classrooms for the recording of diagnostic behavioural data of children with severe behaviour disorders. Using such a system, a teacher takes control of data archiving to document an incident after it has occurred. Regarding BufferWare, it is a system aimed at semi-public spaces in which people can save recorded data, if desired, using a touch-screen. They are also able of viewing previous recordings. In this case, BufferWare was placed in a social area of an academic building. These two systems were proposed in order to conduct an experiment to assess issues of the selective archiving model, such as ownership of data, choice of which data should be saved or deleted, visibility and awareness of recordings, and trust in the fact that policies are being followed and features were implemented as described.

Altcare (Shoaib, Elbrandt, Dragon, & Ostermann, 2010) is a monitoring system based on a static network of video cameras aimed for emergency detection at home. It focuses on fall detection, and works without any manual initialisation or interaction with older persons. When a fall is detected, Altcare first attempts to get the confirmation from the involved person. Afterwards, if the verification is positive, it automatically communicates the emergency to the responsible person. In addition to information concerning the involved person, the system transmits a patch of the video that shows the emergency. In order to protect the privacy, the person is replaced by the silhouette in the video. Furthermore, Altcare enables system administrators to check the state of the person at any time, if so desired by the monitored person.

In Sohn, Plataniotis, and Ro (2010), a privacy-preserving watch list screening system for video-surveillance is proposed. Depending on the group of interest the watch list comprise either black-listed or white-listed identities. Hence, the goal is to verify whether a person is enrolled in the watch list or not. In order to preserve privacy it relies on homomorphic encryption. This system is composed of a network of distributed cameras and a central management server (CMS) which owns the watch list database. This database is considered private and it contains face feature vectors corresponding to identities of interest. In this system, cameras analyse the recorded raw video in order to detect face regions. When a face is detected, it is encrypted and sent to the CMS where face feature vectors are retrieved from the watch list database. Afterwards, a comparison between the face coming from a camera and those from the watch list is performed in the encryption domain. Then a result confirming whether the identity is included in the watch list or not is encrypted and sent back to the camera. Finally, according to the obtained result the camera notifies it to the security service.

TrustCAM (Winkler & Rinner, 2010a, 2010b) is a smart camera that provides security and privacy-protection based on trusted computing. By using this smart camera video bitstreams are digitally signed. Thereby, the manipulation of recorded images or videos can be detected, origin of images are provided, and multi-level access control is supported. Each TrustCAM node is operated by a central control station (CS) which runs a protected database where cryptographic keys generated during camera setup are securely stored. Camera nodes encrypt privacy sensitive image data such as faces or number plates using a special cryptographic key stored in each one. Moreover, an abstracted version of the sensitive regions as, for example, a silhouette can also be generated and encrypted. These encrypted sensitive regions cannot be decrypted without having access to the CS. Indeed, each representation requires a different key in order to decrypt it, thereby providing multi-level access control.

Finally, PAAS (Barhm, Qwasmi, Qureshi, & el Khatib, 2011) is a privacy-aware surveillance system that enables monitored subjects to specify their privacy preferences via gestures. It uses face recognition technology in order to map individuals to their privacy preferences and security credentials that are encoded using an extension of the P3P-APPEL framework (Cranor, Langheinrich, & Marchiori, 2002). Users have access to one of three privacy settings, namely no privacy, blurred face and blurred full body.

5. Discussion

5.1. Recognition of the region of interest

As it has been seen throughout this survey, privacy protection methods mainly focus on preserving visual privacy in images and videos, either modifying completely the whole image or only a region of interest. Although the detection of the region of interest has not been specifically considered in this work, correct detection is fundamental in order to make protection methods work as desired. For instance, it could be imagined a hypothetical case in which a scrambling algorithm was used to obfuscate a person's face in a video. Supposing that the video sampling rate is 25 fps, there are 3000 frames in 2 min of video where a face appears. It could be supposed also that the hit rate of the used face detector was around 98%. Under these assumptions, the subject face will be poorly detected in roughly 60 frames of video, thereby not enabling the obscuring algorithm to work properly in protecting privacy. If the face is fully observable in a single frame, subject identification can already be performed. Not only that, the face could also be inferred observing multiple frames in which the face detector failed. Hence, it is essential that the computer vision techniques (segmentation, object detection, tracking, recognition, etc.) involved before the obscuring process are reliable and robust enough in order to guarantee the effectiveness of the protection method.

5.2. Privacy and intelligibility trade-off

Although we have mentioned the privacy-intelligibility trade-off in the introduction of Section 3.4, let us provide a further discussion here. Zhao and Stasko (1998) compared several image filters (blurring, pixelating, edge-detector, live-shadow and shadow-view) and found out that whereas intelligibility is provided almost in all of the tested filters, actors were also recognised in most of the cases. Boyle et al. (2000) performed an in depth evaluation of blurring and pixelating to analyse how the variation of the filtration level affects to this balance. The obtained results suggested that blurring and, to a lesser extent, pixelating may provide a balance, but it would be a precarious balance at best. Furthermore, Neustaedter et al. (2006) state that blurring by itself does not suffice for privacy protection when a balance is needed. Therefore, it seems that image filters can hardly provide a balance between privacy and information utility, and generally privacy loses in this balance. Nevertheless, when such a balance is not needed, image filters may be used. As reported by Agrawal and Narayanan (2011), blurring is enough to hide the identity if gait is not involved. But gait and other temporal characteristics are difficult to hide if there is some familiarity between the subject and the user. Anyway, it would be interesting to expand these studies to also include visual abstraction methods and all those where such a balance could be analysed.

5.3. Real-time applications

Regarding the techniques that could be used in real-time intelligent monitoring systems, apart from image filtering, the remainder techniques would be impracticable because they are very time consuming. In such systems, all the involved computer vision algorithms along with the protection method must jointly work in real time. Like inpainting techniques, encryption ones have high computational requirements. Nevertheless, some light-weight encryption techniques may work properly. In turn, some image inpainting techniques are designed to work in real time too. The election of which protection method to use depends on the application requirements.

5.4. Privacy model and evaluation

Relying on a formal model to guarantee privacy preservation is required in order to have a common framework in which privacy solutions can be developed and evaluated. Common definitions about what visual privacy should be, when privacy is protected, which are the sensitive areas, and so on are needed. A model that currently fulfils some of these requirements is the face de-identification based on the *k-same* family of algorithms. Although it guarantees that a de-identified face cannot be recognised with a probability higher than $\frac{1}{k}$, inference channels are not considered. In that sense, Saini et al. (2014) propose a model that takes this into account. This model measures the privacy loss of the individuals that usually inhabitant a surveillance area. This measure is given as a continuous variable in the range [0, 1]. As far as we know, this is the only approach that currently measures privacy loss in such a way. Regarding the remainder methods, despite they are not based on a formal model, it can be intuitively realised that they provide different protection levels. For instance, naïve image filtering techniques like blurring or pixelating are not sufficient to protect privacy when intelligibility is involved. But blanking out or encrypting sensitive regions could be enough.

Also related to the previous point is how the evaluation of visual privacy protection is performed. We have not found a common framework for it. Although the aforementioned model for measuring privacy loss is very valuable, it cannot be used to build such a framework that automatically performs the comparison because current technology is not robust enough. This would have been very useful in order to perform an exhaustive comparison covering all of the reviewed methods. Moreover, the lack of more manually-labelled privacy datasets also contributes to deteriorate visual privacy evaluation. The only datasets focused on privacy that have been found in the literature are PEViD-HD (Korshunov & Ebrahimi, 2013) and PEViD-UHD (Korshunov & Ebrahimi, 2014). Both datasets consider video surveillance scenarios and provide high definition video sequences and ultra high definition ones, respectively. A similar dataset for ambient-assisted living would be appreciated where, in addition, other kind of visual sensors were included (*e.g.* RGB-D sensors). In addition to assess the grade in which privacy is protected, to what extent useful information is retained should also be studied.

In the absence of such evaluation mechanisms, several works have been found in which some of the protection methods have been tested by users as part of the experimentation. These experiments are mainly based on conducting interviews and questionnaires with users where their skills in extracting useful information from a privacy-protected image (subject identification, pose, activity, presence, etc.) are evaluated (Korshunov et al., 2012). Results obtained from an objective study are missed.

5.5. User studies about privacy requirements

Concerning users, there is no agreement about privacy requirements because privacy is highly subjective and which information is considered sensitive depends on each individual. Despite of what has been previously said about blurring and pixelating not being effective in providing a balance between privacy and intelligibility, some users that have participated in studies about this matter feel satisfied with these filters when they are configured correctly (Boyle et al., 2000; Lander et al., 2001; Neustaedter et al., 2006; Zhao & Stasko, 1998). Other participants felt that there is not such a balance because as they were able to know what subjects were doing, privacy was not being suitably protected. Indeed, while some of the participants showed interest in using video cameras at work and even at their homes imposing some constraints, others commented that they would not use them because video cameras are very intrusive. Anyway, what can be extracted about this lack of consensus is that privacy is a very subjective topic and user's requirements vary.

Considering which protection methods preserve privacy better for a fall detection system, a user survey was conducted by Edgcomb and Vahid (2012). They analysed several visual representations: blur, silhouette, oval, box, and trailing arrows. Results indicated that silhouettes and blur were perceived to provide insufficient privacy, whereas an oval provides sufficient perceived privacy while still supporting fall detection accuracy of 89%.

Others have also researched about possible features of subjects' sense of security and privacy for video surveillance systems (Babaguchi, Koshimizu, Umata, & Toriyama, 2009; Koshimizu, Toriyama, & Babaguchi, 2006). Results showed how subjects classify viewers that monitor them using cameras in: familiar persons, unfamiliar persons and persons in duty. Moreover, subjects expect a very familiar person to protect them in an emergency. Concerning the disclosure of private information, it is affected by the closeness between subjects and viewers. Familiarity or closeness between persons facilitates the recognition of each other. It is also curious to see how people in their 50s are less sensitive to privacy than those in their 20–40s. Regarding older people, most of them agree with using silhouettes as a visual representation in order to protect their privacy. Furthermore, they show interest in customising the system operation. They demand control over the camera and who has access to the captured or stored information (*e.g.* turning a camera off, setting up the filtration level, seeing how they are being watched, and so on). However, there are others who consider that they do not require a monitoring system because they are independent enough (Demiris, Oliver, Giger, Skubic, & Rantz, 2009).

Finally, it is necessary that users demand protection methods in order to get visual privacy protection techniques widely used in video surveillance systems. Without such a demand, main actors of the security market, more interested in making a profit, will not pay attention to privacy issues because they do not have enough motivation to kick-start self-regulation (Gutwirth, Leenes, De Hert, & Poullet, 2012). We consider that a strong demand of privacy-aware systems will increase research on this field also benefiting to future AAL systems. Either way, more discussion is needed in visual privacy, privacy requirements of users according to applications need to be collected, and more research on visual privacy protection is required.

6. Conclusion

As we have seen throughout this work, the proliferation of networks of video cameras for surveillance in public spaces and the advances in computer vision have led to a situation in which the privacy of individuals is being compromised. Moreover, nowadays video cameras are being used more often in ambient-assisted living applications that operate in private spaces. Because of this, expert and intelligent systems that handle these tasks should take privacy into account by means of new tools that restore the individual's right to privacy.

In this paper, an up-to-date review of visual privacy protection methods has been provided. As far as we know, this is the first review that focuses on the protection methods and tries to give a comprehensive classification of all of them. In our classification, we have considered the following categories: intervention, blind vision, secure processing, redaction and data hiding. As it has been seen, redaction methods cover the largest part of this work and they have also been classified according to the way in which imagery data is modified: image filtering, encryption, face de-identification, object removal and visual abstraction.

Given that the previous methods are required for expert and intelligent systems for video surveillance and ambient-assisted living, some of them have also been reviewed. Such systems are developed as a framework to provide some kind of privacy protection. However, only using some of the described methods is not enough to build a privacy-aware system. As it has been seen, they have to support also other mechanisms to implement privacy policies. These policies specify what data is protected, how it is protected, who has access, and a variety of other fundamental aspects concerning involved subjects, ongoing activities, captured data, and so on. Despite of visual privacy protection being necessary in expert and intelligent systems for video surveillance, due to its increasing power of tracking, storing and indexing of our public daily activities, the use of privacy protection methods can also be justified for such systems but from an ambient-assisted living perspective. In that sense, privacy-aware smart cameras are essential in the construction of future low-cost systems for telecare applications aimed at older and disabled people.

As another contribution of this work, we have also provided a discussion about important factors and current limitations of visual privacy protection. For instance, a key aspect, *i.e.* the correct detection of the region of interest, has been highlighted. It is a very relevant topic because it can be considered as the first building block of privacy protection and, therefore, it would deserve a self-standing review of the related research fields. We have also mentioned the problem faced by redaction methods, *i.e.* the privacy and intelligibility trade-off, and the necessity of expanding these studies to cover other protection methods. Furthermore, we have stood out the lack of a formal model and evaluation mechanisms that would enable to make fair comparisons between different protection methods and, more importantly, it would provide guarantees about their accuracy in protecting privacy.

This work provides a comprehensive picture of how to protect visual privacy, so it can be a good starting point for novel researchers in the field of expert and intelligent systems, and more specifically in intelligent monitoring systems for video surveillance and ambient-assisted living. For future research directions we consider that there are two main axis that need special attention. On the one hand, as aforementioned, a standard way to quantify and evaluate visual privacy protection is needed so as to fairly compare works in this matter. It would be better if this is included as a part of a formal model that considers more privacy related aspects. On the other hand, recognition accuracy of sensitive regions needs to be improved, so research is required in more robust computer vision algorithms for recognition, tracking and event detection.

Regarding other future research directions, protection methods would also deserve some consideration. Although a lot of work has been done so far, novel redaction methods that provide a better balance between privacy and intelligibility are welcome. For instance, a textual representation of the scene, where the essential context (*e.g.* individuals, events, etc.) is captured, could be enough for some applications like telecare. Privacy preferences are also very important. Mechanisms to empower users and put them in control of their private data captured by intelligent systems should be researched. It would be appreciated to have a common way to let users specify their preferences so they can decide who, how, where and when they are watched in normal scenarios where no law infringement is involved. Finally, users should be able of knowing and tracking which systems have collected data about them in order to enable them to take legal actions in this matter just in case they need it. Therefore, work in these areas may lead to obtain more effective and reliable visual privacy protection methods and systems in a near future and also helps to increase the users' acceptance of using video cameras in private spaces.

Acknowledgement

This work has been partially supported by the Spanish Ministry of Science and Innovation under project "Sistema de visión para la monitorización de la actividad de la vida diaria en el hogar" (TIN2010-20510-C04-02) and by the European Commission under project "caring4U – A study on people activity in private spaces: towards a multisensor network that meets privacy requirements" (PIEF-GA-2010-274649). José Ramón Padilla López and Alexandros Andre Chaaraoui acknowledge financial support by the Conselleria d'Educació, Formació i Ocupació of the Generalitat Valenciana (fellowship ACIF/2012/064 and ACIF/2011/160 respectively). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Abraham, A. R., Prabhavathy, K. A., & Shree, D. J. (2012). A survey on video inpainting. *International Journal of Computer Applications, 56*, 43–47.

Adams, A. A., & Ferryman, J. M. (2013). The future of video analytics for surveillance and its ethical implications. *Security Journal*. http://dx.doi.org/10.1057/sj.2012.48.

Agi, I., & Gong, L. (1996). An empirical study of secure MPEG video transmissions. In *Proceedings of the symposium on network and distributed system security* (pp. 137–144). IEEE.

Agrawal, P., & Narayanan, P. (2011). Person de-identification in videos. *IEEE Transactions on Circuits and Systems for Video Technology, 21*, 299–310.

Augusto, J., Nakashima, H., & Aghajan, H. (2010). Ambient intelligence and smart environments: A state of the art. In H. Nakashima, H. Aghajan, & J. Augusto (Eds.), *Handbook of ambient intelligence and smart environments* (pp. 3–31). US: Springer.

Avidan, S., & Butman, M. (2006). Blind vision. In A. Leonardis, H. Bischof, & A. Pinz (Eds.), *Computer vision – ECCV 2006. Lecture notes in computer science* (Vol. 3953, pp. 1–13). Berlin, Heidelberg: Springer.

Avidan, S., Elbaz, A., Malkin, T., & Moriarty, R. (2009). Oblivious image matching. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 49–64). London: Springer.

Baaziz, N., Lolo, N., Padilla, O., & Petngang, F. (2007). Security and privacy protection for automated video surveillance. In *2007 IEEE international symposium on signal processing and information technology* (pp. 17–22).

Babaguchi, N., Koshimizu, T., Umata, I., & Toriyama, T. (2009). Psychological study for designing privacy protected video surveillance system: PriSurv. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 147–164). London: Springer.

Baran, I., & Popović, J. (2007). Automatic rigging and animation of 3D characters. *ACM Transactions on Graphics, 26*. http://dx.doi.org/10.1145/1276377.1276467.

Barhm, M., Qwasmi, N., Qureshi, F., & el Khatib, K. (2011). Negotiating privacy preferences in video surveillance systems. In K. Mehrotra, C. Mohan, J. Oh, P. Varshney, & M. Ali (Eds.), *Modern approaches in applied intelligence. Lecture notes in computer science* (Vol. 6704, pp. 511–521). Berlin, Heidelberg: Springer.

Bertalmio, M., Sapiro, G., Caselles, V., & Ballester, C. (2006). Image inpainting. In *Computer graphics and interactive techniques. Proceedings of the 27th annual conference on SIGGRAPH '00* (pp. 417–424). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co.

Bitouk, D., Kumar, N., Dhillon, S., Belhumeur, P., & Nayar, S. K. (2008). Face swapping: Automatically replacing faces in photographs. *ACM Transactions on Graphics, 27*, 39:1–39:8.

Borosán, P., Jin, M., DeCarlo, D., Gingold, Y., & Nealen, A. (2012). RigMesh: Automatic rigging for part-based shape modeling and deformation. *ACM Transactions on Graphics, 31*, 198:1–198:9.

Boult, T. (2005). PICO: Privacy through invertible cryptographic obscuration. In *Computer vision for interactive and intelligent environment* (pp. 27–38). IEEE.

Boyle, M., Edwards, C., & Greenberg, S. (2000). The effects of filtered video on awareness and privacy. In *Computer supported cooperative work, Proceedings of the 2000 ACM conference on CSCW '00* (pp. 1–10). New York, NY, USA: ACM.

Caloyannides, M. A. (2003). Society cannot function without privacy. *IEEE Security and Privacy, 1*, 84–86.

Candes, E., Romberg, J., & Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory, 52*, 489–509.

Cardinaux, F., Bhowmik, D., Abhayaratne, C., & Hawley, M. S. (2011). Video based technology for ambient assisted living: A review of the literature. *Journal of Ambient Intelligence and Smart Environments, 3*, 253–269.

Carrillo, P., Kalva, H., & Magliveras, S. (2010). Compression independent reversible encryption for privacy in video surveillance. *EURASIP Journal on Information Security, 2009*, 1–13. http://dx.doi.org/10.1155/2009/429581. Article ID: 2009:429581.

Cavoukian, A. (2013). Surveillance, Then and Now: Securing Privacy in Public Spaces. Technical Report Information and Privacy Commmissioner of Ontario.

Chaaraoui, A. A., Climent-Pérez, P., & Flórez-Revuelta, F. (2012a). An efficient approach for multi-view human action recognition based on bag-of-key-poses. In A. A. Salah, J. Ruiz-del Solar, c. Meriçli, & P.-Y. Oudeyer (Eds.), *Human behavior understanding. Lecture notes in computer science* (Vol. 7559, pp. 29–40). Berlin, Heidelberg: Springer.

Chaaraoui, A. A., Climent-Pérez, P., & Flórez-Revuelta, F. (2012b). A review on vision techniques applied to human behaviour analysis for ambient-assisted living. *Expert Systems with Applications, 39*, 10873–10888.

Chan, C. S., & Liu, H. (2009). GMM-QNT hybrid framework for vision-based human motion analysis. In *IEEE international conference on fuzzy systems, 2009. FUZZ-IEEE 2009* (pp. 1820–1825).

Chang, Y., Yan, R., Chen, D., & Yang, J. (2006). People identification with limited labels in privacy-protected video. In *2006 IEEE international conference on multimedia and expo* (pp. 1005–1008).

Chan, T., Shen, J., & Zhou, H.-M. (2006). Total variation wavelet inpainting. *Journal of Mathematical Imaging and Vision, 25*, 107–125.

Chattopadhyay, A., & Boult, T. (2007). PrivacyCam: A privacy preserving camera using uCLinux on the blackfin DSP. In *IEEE conference on computer vision and pattern recognition, 2007. CVPR '07* (pp. 1–8).

Chen, D., Chang, Y., Yan, R., & Yang, J. (2007). Tools for protecting the privacy of specific individuals in video. *EURASIP Journal on Advances in Signal Processing, 2007*, 1–9. http://dx.doi.org/10.1155/2007/75427. Article ID: 2007:075427.

Chen, D., Chang, Y., Yan, R., & Yang, J. (2009). Protecting personal identification in video. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 115–128). London: Springer.

Chen, Y.-Y., Cho, C.-W., Lin, S.-H., Lai, H.-Y., Lo, Y.-C., Chen, S.-Y., et al. (2012). A vision-based regression model to evaluate parkinsonian gait from monocular image sequences. *Expert Systems with Applications, 39*, 520–526.

Chen, Y., Dang, G., Cheng, Z., & Xu, K. (2014). Fast capture of personalized avatar using two kinects. *Journal of Manufacturing Systems, 33*, 233–240.

Cheung, S. -C. S., Zhao, J., & Venkatesh, M. (2006). Efficient object-based video inpainting. In *2006 IEEE international conference on image processing* (pp. 705–708).

Cheung, S. -C., Paruchuri, J., & Nguyen, T. (2008). Managing privacy data in pervasive camera networks. In *15th IEEE international conference on image processing, 2008. ICIP 2008* (pp. 1676–1679).

Cheung, S.-C., Venkatesh, M., Paruchuri, J., Zhao, J., & Nguyen, T. (2009). Protecting and managing privacy information in video surveillance systems. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 11–33). London: Springer.

Chinomi, K., Nitta, N., Ito, Y., & Babaguchi, N. (2008). PriSurv: Privacy protected video surveillance system using adaptive visual abstraction. In S. Satoh, F. Nack, & M. Etoh (Eds.), *Advances in multimedia modeling. Lecture notes in computer science* (Vol. 4903, pp. 144–154). Berlin, Heidelberg: Springer.

Cook, D. J., Augusto, J. C., & Jakkula, V. R. (2009). Ambient intelligence: Technologies, applications, and opportunities. *Pervasive and Mobile Computing, 5*, 277–298.

Cootes, T., Edwards, G., & Taylor, C. (1998). Active appearance models. In H. Burkhardt & B. Neumann (Eds.), *Computer vision – ECCV'98. Lecture notes in computer science* (Vol. 1407, pp. 484–498). Berlin He idelberg: Springer.

Cootes, T., Edwards, G., & Taylor, C. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 23*, 681–685.

Coudert, F. (2010). When video cameras watch and screen: Privacy implications of pattern recognition technologies. *Computer Law & Security Review, 26*, 377–384.

Coutaz, J., Bérard, F., Carraux, E., Astier, W., & Crowley, J. L. (1999). CoMedi: Using computer vision to support awareness and privacy in mediaspaces. In *CHI '99 extended abstracts on human factors in computing systems CHI EA '99* (pp. 13–14). New York, NY, USA: ACM.

Cox, I., Honsinger, C., Miller, M., & Bloom, J. (2002). Digital watermarking. *Journal of Electronic Imaging, 11*. 414–414.

Cranor, L., Langheinrich, M., & Marchiori, M. (2002). A P3P preference exchange language 1.0 (APPEL1.0). <http://www.w3.org/TR/P3P-preferences/> Accessed 08.01.2015.

Criminisi, A., Perez, P., & Toyama, K. (2003). Object removal by exemplar-based inpainting. In *Proceedings of IEEE computer society conference on computer vision and pattern recognition* (Vol. 2, pp. II-721–II-728).

Criminisi, A., Perez, P., & Toyama, K. (2004). Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing, 13*, 1200–1212.

Dee, H., & Velastin, S. (2008). How close are we to solving the problem of automated visual surveillance? *Machine Vision and Applications, 19*, 329–343.

De la Hunty, M., Asthana, A., & Goecke, R. (2010). Linear facial expression transfer with active appearance models. In *2010 20th International conference on pattern recognition (ICPR)* (pp. 3789–3792).

Demiris, G., Oliver, D., Giger, J., Skubic, M., & Rantz, M. (2009). Older adults' privacy considerations for vision based recognition methods of eldercare applications. *Technology and Health Care, 17*, 41–48.

Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory, 52*, 1289–1306.

Dufaux, F. (2011). Video scrambling for privacy protection in video surveillance: Recent results and validation framework. In *Mobile Multimedia/Image Processing, Security, and Applications 2011, Proceedings of SPIE* (Vol. 8063, pp. 1–12). http://dx.doi.org/10.1117/12.883948 Article ID: 806302.

Dufaux, F., & Ebrahimi, T. (2006). Scrambling for video surveillance with privacy. In *Conference on computer vision and pattern recognition workshop, 2006. CVPRW '06* (p. 160). http://dx.doi.org/10.1109/CVPRW.2006.184.

Dufaux, F., & Ebrahimi, T. (2008). H.264/AVC video scrambling for privacy protection. In *15th IEEE international conference on image processing, 2008. ICIP 2008* (pp. 1688–1691).

Dufaux, F., & Ebrahimi, T. (2010). A framework for the validation of privacy protection solutions in video surveillance. In *2010 IEEE international conference on multimedia and expo (ICME)* (pp. 66–71).

Dufaux, F., & Ebrahimi, T. (2008b). Scrambling for privacy protection in video surveillance systems. *IEEE Transactions on Circuits and Systems for Video Technology, 18*, 1168–1174.

Dufaux, F., Ouaret, M., Abdeljaoued, Y., Navarro, A., Vergnenègre, F., & Ebrahimi, T. (2006). Privacy enabling technology for video surveillance. In *Mobile multimedia/image processing for military and security applications, proceedings of SPIE* (Vol. 6250, pp. 1–12). http://dx.doi.org/10.1117/12.664945 Article ID: 62500M.

EC (2008). Seniorwatch 2. Assessment of the Senior Market for ICT Progress and Developments. Technical Report European Commission.

EC (2010). Overview of the European strategy in ICT for Ageing Well. Technical Report European Commission.

Edgcomb, A., & Vahid, F. (2012). Privacy perception and fall detection accuracy for in-home video assistive monitoring with privacy enhancements. *ACM SIGHIT Record, 2*, 6–15.

Efros, A., & Leung, T. (1999). Texture synthesis by non-parametric sampling. In *The proceedings of the seventh IEEE international conference on computer vision, 1999* (Vol. 2, pp. 1033–1038).

Efros, A. A., & Freeman, W. T. (2001). Image quilting for texture synthesis and transfer. In *Computer graphics and interactive techniques. Proceedings of the 28th annual conference on SIGGRAPH '01* (pp. 341–346). New York, NY, USA: ACM.

Erkin, Z., Franz, M., Guajardo, J., Katzenbeisser, S., Lagendijk, I., & Toft, T. (2009). Privacy-preserving face recognition. In I. Goldberg & M. Atallah (Eds.), *Privacy enhancing technologies. Lecture notes in computer science* (Vol. 5672, pp. 235–253). Berlin, Heidelberg: Springer.

Erturk, S. (2007). Multiplication-free one-bit transform for low-complexity block-based motion estimation. *Signal Processing Letters, IEEE, 14*, 109–112.

Fan, J., Luo, H., Hacid, M.-S., & Bertino, E. (2005). A novel approach for privacy-preserving video sharing. In *Information and knowledge management, proceedings of the fourteenth ACM international conference on CIKM '05* (pp. 609–616). New York, NY, USA: ACM.

Fidaleo, D. A., Nguyen, H.-A., & Trivedi, M. (2004). The networked sensor tapestry (NeST): A privacy enhanced software architecture for interactive analysis of data in video-sensor networks. In *Video surveillance & sensor networks, proceedings of the ACM second international workshop on VSSN '04* (pp. 46–53). New York, NY, USA: ACM.

Fleck, S., & Strasser, W. (2008). Smart camera based monitoring system and its application to assisted living. *Proceedings of the IEEE, 96*, 1698–1714.

Frome, A., Cheung, G., Abdulkader, A., Zennaro, M., Wu, B., Bissacco, A., Adam, H., Neven, H., & Vincent, L. (2009). Large-scale privacy protection in google street view. In *2009 IEEE 12th international conference on computer vision* (pp. 2373–2380).

Gerrard, G., & Thompson, R. (2011). Two million cameras in the UK. *CCTV Image, 42*, 10–12.

Ghanbari, A., & Soryani, M. (2011). Contour-based video inpainting. In *2011 Seventh Iranian machine vision and image processing (MVIP)* (pp. 1–5). IEEE.

Goessl, L. (2012). New japanese security camera scans 36 million faces per second. URL http://digitaljournal.com/article/321848 Accessed 08.05.2014.

Granados, M., Tompkin, J., Kim, K., Grau, O., Kautz, J., & Theobalt, C. (2012). How not to be seen – object removal from videos of crowded scenes. *Computer Graphics Forum, 31*, 219–228.

Gross, R., Airoldi, E., Malin, B., & Sweeney, L. (2006). Integrating utility into face de-identification. In G. Danezis & D. Martin (Eds.), *Privacy enhancing technologies. Lecture notes in computer science* (Vol. 3856, pp. 227–242). Berlin, Heidelberg: Springer.

Gross, R., Sweeney, L., de la Torre, F., & Baker, S. (2006). Model-based face de-identification. In *Conference on computer vision and pattern recognition workshop, 2006. CVPRW '06* (pp. 161–168).

Gross, R., Sweeney, L., Cohn, J., Torre, F., & Baker, S. (2009). Face de-identification. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 129–146). London: Springer.

Guangzhen, L., Yoshimichi, I., Xiaoyi, Y., Naoko, N., & Noboru, B. (2010). Recoverable privacy protection for video content distribution. *EURASIP Journal on Information Security, 200*, 1–11. http://dx.doi.org/10.1155/2009/293031. Article ID: 2009:293031.

Gutwirth, S., Leenes, R., De Hert, P., & Poullet, Y. (2012). *European data protection: In good health?* Springer.

Harvey, A. (2010). Camoflash – anti-paparazzi clutch. URL http://ahprojects.com/projects/camoflash/ Accessed 08.05.2014.

Hayes, G., & Truong, K. (2009). Selective archiving: A model for privacy sensitive capture and access technologies. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 165–184). London: Springer.

He, L., Bleyer, M., & Gelautz, M. (2011). Object removal by depth-guided inpainting. In *The Austrian association for pattern Recognition (OAGM/AAPR) workshop 2011 (OAGM2011)* (pp. 1–8).

Hodgins, J. K., O'Brien, J. F., & Tumblin, J. (1998). Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics, 4*, 307–316.

Hogue, A., Gill, S., & Jenkin, M. (2007). Automated avatar creation for 3D games. In *Future play, proceedings of the 2007 conference on future play '07* (pp. 174–180). New York, NY, USA: ACM.

Hudson, S. E., & Smith, I. (1996). Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems. In *Computer supported cooperative work, proceedings of the 1996 ACM conference on CSCW '96* (pp. 248–257). New York, NY, USA: ACM.

Igehy, H., & Pereira, L. (1997). Image replacement through texture synthesis. In *Proceedings of the international conference on image processing* (Vol. 3, pp. 186–189).

Ito, I., & Kiya, H. (2009). One-time key based phase scrambling for phase-only correlation between visually protected images. *EURASIP Journal on Information Security, 2009*, 1–11. http://dx.doi.org/10.1155/2009/841045. Article ID: 2009:841045.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*, 201–211.

Kitahara, I., Kogure, K., & Hagita, N. (2004). Stealth vision for protecting privacy. In *Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004* (Vol. 4, pp. 404–407).

Kokaram, A., Morris, R., Fitzgerald, W., & Rayner, P. (1995). Interpolation of missing data in image sequences. *IEEE Transactions on Image Processing, 4*, 1509–1519.

Koochari, A., & Soryani, M. (2010). Exemplar-based video inpainting with large patches. *Journal of Zhejiang University Science C, 11*, 270–277.

Korshunov, P., & Ebrahimi, T. (2013). PEViD: Privacy evaluation video dataset. In *Applications of Digital Image Processing XXXVI, Proceedings of SPIE* (Vol. 8856, pp. 1–9). http://dx.doi.org/10.1117/12.2030974 Article ID: 88561S.

Korshunov, P., & Ebrahimi, T. (2014). UHD video dataset for evaluation of privacy. In *Sixth international workshop on quality of multimedia experience (QoMEX 2014)*.

Korshunov, P., Araimo, C., De Simone, F., Velardo, C., Dugelay, J., & Ebrahimi, T. (2012). Subjective study of privacy filters in video surveillance. In *2012 IEEE 14th international workshop on multimedia signal processing (MMSP)* (pp. 378–382).

Koshimizu, T., Toriyama, T., & Babaguchi, N. (2006). Factors on the sense of privacy in video surveillance. In *Continuous archival and retrieval of personal experiences, proceedings of the 3rd ACM workshop on CARPE '06* (pp. 35–44). New York, NY, USA: ACM.

Lander, K., Bruce, V., & Hill, H. (2001). Evaluating the effectiveness of pixelation and blurring on masking the identity of familiar faces. *Applied Cognitive Psychology, 15*, 101–116.

Langheinrich, M. (2001). Privacy by design – principles of privacy-aware ubiquitous systems. In *Ubiquitous Computing, proceedings of the 3rd international conference on UbiComp '01* (pp. 273–291). Springer-Verlag.

Lin, Y., Wang, S., Lin, Q., & Tang, F. (2012). Face swapping under large pose variations: A 3D model based approach. In *2012 IEEE international conference on multimedia and expo (ICME)* (pp. 333–338).

Li, S., Zhao, Y., Qu, B., & Wang, J. (2013). Image scrambling based on chaotic sequences and veginère cipher. *Multimedia Tools and Applications, 66*, 573–588.

Lyons, M., Plante, A., Jehan, S., Inoue, S., & Akamatsu, S. (1998). Avatar creation using automatic face processing. In *Proceedings of the sixth ACM international conference on multimedia* (pp. 427–434). ACM.

Ma, W., & Ma, Q. (2011). An effective method for removing large object. In *2011 4th international congress on image and signal processing (CISP)* (Vol. 1, pp. 80–84).

Macq, B., & Quisquater, J.-J. (1995). Cryptology for digital tv broadcasting. *Proceedings of the IEEE, 83*, 944–957.

Martínez-Ponte, I., Desurmont, X., Meessen, J., & Delaigle, J. -F. (2005). Robust human face hiding ensuring privacy. In *Proceedings of the international workshop on image analysis for multimedia interactive services (WIAMIS)* (Vol. 4).

Masnou, S., & Morel, J. -M. (1998). Level lines based disocclusion. In *1998 international conference onimage processing, 1998. ICIP 98. Proceedings* (Vol. 3, pp. 259–263).

Mitskog, T. F. R., & Ralston, R. A. (2012). Camera blocker for a device with an integrated camera that uses a thin film organic polymer.

Morris, M. E., Adair, B., Miller, K., Ozanne, E., Hansen, R., Pearce, A. J., et al. (2013). Smart-home technologies to assist older people to live well at home. *Journal of Aging Science, 1*. http://dx.doi.org/10.4172/2329-8847.1000101.

Morse, B., & Schwartzwald, D. (1998). Isophote-based interpolation. In *1998 International conference on image processing, 1998. ICIP 98. Proceedings* (Vol. 3, pp. 227–231).

Neustaedter, C., & Greenberg, S. (2003). The design of a context-aware home media space for balancing privacy and awareness. In A. Dey, A. Schmidt, & J. McCarthy (Eds.), *UbiComp 2003: Ubiquitous computing. Lecture notes in computer science* (Vol. 2864, pp. 297–314). Berlin, Heidelberg: Springer.

Neustaedter, C., Greenberg, S., & Boyle, M. (2006). Blur filtration fails to preserve privacy for home-based video conferencing. *ACM Transactions on Computer–Human Interaction, 13*, 1–36.

Newton, E., Sweeney, L., & Malin, B. (2005). Preserving privacy by de-identifying face images. *IEEE Transactions on Knowledge and Data Engineering, 17*, 232–243.

Ng, P. L., minn Ang, L., & Seng, K. P. (2010). Privacy preserving stereoscopic vision with one-bit transform. In *2010 third IEEE international conference on computer science and information technology (ICCSIT)* (Vol. 9, pp. 70–74).

Ni, Z., Shi, Y.-Q., Ansari, N., & Su, W. (2006). Reversible data hiding. *IEEE Transactions on Circuits and Systems for Video Technology, 16*, 354–362.

O'Brien, A., & Mac Ruairi, R. (2009). Survey of assistive technology devices and applications for aging in place. In *Second international conference on advances in human-oriented and personalized mechanisms, technologies, and services, 2009. CENTRIC '09* (pp. 7–12).

Olivieri, D. N., Conde, I. G., & Sobrino, X. A. V. (2012). Eigenspace-based fall detection and activity recognition from motion templates and machine learning. *Expert Systems with Applications, 39*, 5935–5945.

Pande, A., & Zambreno, J. (2013). Securing multimedia content using joint compression and encryption. In *Embedded multimedia security systems* (pp. 23–30). London: Springer.

Park, S., & Kautz, H. (2008). Privacy-preserving recognition of activities in daily living from multi-view silhouettes and RFID-based training. In *AAAI fall symposium on AI in Eldercare: New solutions to old problems*.

Paruchuri, J., & Cheung, S. -C. (2008). Joint optimization of data hiding and video compression. In *IEEE international symposium on circuits and systems, 2008. ISCAS 2008* (pp. 3021–3024).

Paruchuri, J. K., Cheung, S.-c. S., & Hail, M. W. (2009). Video data hiding for managing privacy information in surveillance systems. *EURASIP Journal on Information Security, 2009*, 1–18. http://dx.doi.org/10.1155/2009/236139. Article ID: 2009:236139.

Patel, S. N., Summet, J. W., & Truong, K. N. (2009). Blindspot: Creating capture-resistant spaces. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 185–201). London: Springer.

Patwardhan, K., Sapiro, G., & Bertalmio, M. (2007). Video inpainting under constrained camera motion. *IEEE Transactions on Image Processing, 16*, 545–553.

Petitcolas, F., Anderson, R., & Kuhn, M. (1999). Information hiding-a survey. *Proceedings of the IEEE, 87*, 1062–1078.

Qiao, L., Nahrstedt, K., et al. (1997). A new algorithm for MPEG video encryption. In *Proceedings of first international conference on imaging science, systems and technology CISST'97* (pp. 21–29).

Qureshi, F. (2009). Object-video streams for preserving privacy in video surveillance. In *Sixth IEEE international conference on advanced video and signal based surveillance, 2009. AVSS '09* (pp. 442–447).

Ra, M. -R., Govindan, R., & Ortega, A. (2013). P3: Toward privacy-preserving photo sharing. In *Proceedings of the 10th USENIX conference on networked systems design and implementation* (pp. 515–528).

Raju, C., Umadevi, G., Srinathan, K., & Jawahar, C. (2008). Fast and secure real-time video encryption. In *Sixth Indian conference on computer vision, graphics image processing, 2008. ICVGIP '08* (pp. 257–264). IEEE.

Sadeghi, A.-R., Schneider, T., & Wehrenberg, I. (2010). Efficient privacy-preserving face recognition. In D. Lee & S. Hong (Eds.), *Information, security and cryptology – ICISC 2009. Lecture notes in computer science* (Vol. 5984, pp. 229–244). Berlin, Heidelberg: Springer.

Sadimon, S., Sunar, M., Mohamad, D., & Haron, H. (2010). Computer generated caricature: A survey. In *2010 International conference on cyberworlds (CW)* (pp. 383–390).

Saini, M., Atrey, P., Mehrotra, S., & Kankanhalli, M. (2014). W3-privacy: Understanding what, when, and where inference channels in multi-camera surveillance video. *Multimedia Tools and Applications, 68*, 135–158.

Schaar, P. (2010). Privacy by design. *Identity in the Information Society, 3*, 267–274.

Schiff, J., Meingast, M., Mulligan, D. K., Sastry, S., & Goldberg, K. (2009). Respectful cameras: Detecting visual markers in real-time to address privacy concerns. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 65–89). London: Springer.

Senior, A. (2009). Privacy protection in a video surveillance system. In A. Senior (Ed.), *Protecting privacy in video surveillance* (pp. 35–47). London: Springer.

Senior, A., Pankanti, S., Hampapur, A., Brown, L., li Tian, Y., & Ekin, A. (2003). Blinkering Surveillance: Enabling Video Privacy through Computer Vision. Technical Report IBM Research Division.

Senior, A., & Pankanti, S. (2011). Privacy protection and face recognition. In S. Z. Li & A. K. Jain (Eds.), *Handbook of face recognition* (pp. 671–691). London: Springer.

Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y.-L., Ekin, A., et al. (2005). Enabling video privacy through computer vision. *Security Privacy, IEEE, 3*, 50–57.

Shashanka, M. (2010). A privacy preserving framework for gaussian mixture models. In *2010 IEEE international conference on data mining workshops (ICDMW)* (pp. 499–506).

Shashank, J., Kowshik, P., Srinathan, K., & Jawahar, C. (2008). Private content based image retrieval. In *IEEE conference on computer vision and pattern recognition, 2008. CVPR 2008* (pp. 1–8).

Shiratori, T., Matsushita, Y., Tang, X., & Kang, S. B. (2006). Video completion by motion field transfer. In *2006 IEEE computer society conference on computer vision and pattern recognition* (Vol. 1, pp. 411–418).

Shoaib, M., Elbrandt, T., Dragon, R., & Ostermann, J. (2010). Altcare: Safe living for elderly people. In *2010 4th international conference on pervasive computing technologies for healthcare (PervasiveHealth)* (pp. 1–4).

Sohn, H., De Neve, W., & Ro, Y. M. (2011). Privacy protection in video surveillance systems: Analysis of subband-adaptive scrambling in JPEG XR. *IEEE Transactions on Circuits and Systems for Video Technology, 21*, 170–177.

Sohn, H., Plataniotis, K., & Ro, Y. (2010). Privacy-preserving watch list screening in video surveillance system. In G. Qiu, K. Lam, H. Kiya, X.-Y. Xue, C.-C. Kuo, & M. Lew (Eds.), *Advances in multimedia information processing – PCM 2010. Lecture notes in computer science* (Vol. 6297, pp. 622–632). Berlin, Heidelberg: Springer.

Spanos, G. A., & Maples, T. B. (1995). Performance study of a selective encryption scheme for the security of networked, real-time video. In *International conference on computer communications and networks*.

Srinivasan, S., Tu, C., Regunathan, S., & Sullivan, G. (2007). HD photo: A new image coding technology for digital photography. In *Applications of digital image processing XXX, proceedings of SPIE* (Vol. 6696, pp. 1–19). http://dx.doi.org/10.1117/12.767840 Article ID: 66960A.

Sweeney, L. (2002). -anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10*, 557–570.

Tang, L. (1996). Methods for encrypting and decrypting MPEG video data efficiently. In *Proceedings of the fourth ACM international conference on multimedia* (pp. 219–229). ACM.

Tansuriyavong, S., & Hanaki, S.-i. (2001). Privacy protection by concealing persons in circumstantial video image. In *Proceedings of the 2001 workshop on perceptive user interfaces* (pp. 1–4). ACM.

Tian, Y.-l., Brown, L., Hampapur, A., Lu, M., Senior, A., & Shu, C.-f. (2008). IBM smart surveillance system (S3): Event based video surveillance system with an open and extensible framework. *Machine Vision and Applications, 19*, 315–327.

Tong, L., Dai, F., Zhang, Y., & Li, J. (2010). Prediction restricted H.264/AVC video scrambling for privacy protection. *Electronics Letters, 46*, 47–49.

Tong, L., Dai, F., Zhang, Y., Li, J., & Zhang, D. (2011). Compressive sensing based video scrambling for privacy protection. In *Visual communications and image processing (VCIP)* (pp. 1–4). IEEE.

Turk, M., & Pentland, A. (1991). Face recognition using eigenfaces. In *IEEE computer society conference on computer vision and pattern recognition, 1991. Proceedings CVPR '91* (pp. 586–591).

Upmanyu, M., Namboodiri, A., Srinathan, K., & Jawahar, C. (2009). Efficient privacy preserving video surveillance. In *2009 IEEE 12th international conference on computer vision* (pp. 1639–1646).

Vijay Venkatesh, M., Cheung, S.-c. S., & Zhao, J. (2009). Efficient object-based video inpainting. *Pattern Recognition Letters, 30*, 168–179.

Wagstaff, J. (2004). Using bluetooth to disable camera phones. http://www.loosewireblog.com/2004/09/using_bluetooth.html Accessed 08.05.2014.

Wallace, E., & Diffley, C. (1988). CCTV control rooms ergonomics. Technical Report 14/98 Police Scientific Development Branch (PSDB) UK Home Office.

Wexler, Y., Shechtman, E., & Irani, M. (2007). Space-time completion of video. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*, 463–476.

Whyte, O., Sivic, J., & Zisserman, A. (2009). Get out of my picture! internet-based inpainting. In *Proceedings of the 20th British machine vision conference, London*.

Williams, A., Xie, D., Ou, S., Grupen, R., Hanson, A., & Riseman, E. (2006). Distributed Smart Cameras for Aging in Place. Technical Report DTIC Documen.

Winkler, T., & Rinner, B. (2010b). TrustCAM: Security and privacy-protection for an embedded smart camera based on trusted computing. In *2010 Seventh IEEE international conference on advanced video and signal based surveillance (AVSS)* (pp. 593–600).

Winkler, T., & Rinner, B. (2010a). Securing embedded smart cameras with trusted computing. *EURASIP Journal on Wireless Communications and Networking, 2011*, 1–20. http://dx.doi.org/10.1155/2011/530354. Article ID: 2011:530354.

Winkler, T., & Rinner, B. (2014). Security and privacy protection in visual sensor networks: A survey. *ACM Computing Surveys, 47*, 1–42.

Wu, M. (2001). *Multimedia data hiding*. (Ph.D. thesis). Princeton University.

Xiangdong, L., Junxing, Z., Jinhai, Z., & Xiqin, H. (2008). Image scrambling algorithm based on chaos theory and sorting transformation. *IJCSNS International Journal of Computer Science and Network Security, 8*, 64–68.

Yabuta, K., Kitazawa, H., & Tanaka, T. (2006). A new concept of real-time security camera monitoring with privacy protection by masking moving objects. In *Proceedings of SPIE* (Vol. 6063, pp. 188–199).

Yabuta, K., Kitazawa, H., & Tanaka, T. (2005). A new concept of security camera monitoring with privacy protection by masking moving objects. In Y.-S. Ho & H. Kim (Eds.), *Advances in multimedia information processing – PCM 2005. Lecture notes in computer science* (Vol. 3767, pp. 831–842). Berlin, Heidelberg: Springer.

Yang, M., Bourbakis, N., & Li, S. (2004). Data-image-video encryption. *Potentials, IEEE, 23*, 28–34.

Yu, X., & Babaguchi, N. (2007). Privacy preserving: Hiding a face in a face. In Y. Yagi, S. Kang, I. Kweon, & H. Zha (Eds.), *Computer vision – ACCV 2007. Lecture notes in computer science* (Vol. 4844, pp. 651–661). Berlin, Heidelberg: Springer.

Yu, X., Chinomi, K., Koshimizu, T., Nitta, N., Ito, Y., & Babaguchi, N. (2008). Privacy protecting visual processing for secure video surveillance. In *15th IEEE international conference on image processing, 2008. ICIP 2008* (pp. 1672–1675).

Zeng, W., & Lei, S. (2003). Efficient frequency domain selective scrambling of digital video. *IEEE Transactions on Multimedia, 5*, 118–129.

Zhang, Y., Xiao, J., & Shah, M. (2005). Motion layer based object removal in videos. In *2005 Seventh IEEE workshops on application of computer vision WACV/MOTIONS '05* (Vol. 1, pp. 516–521).

Zhang, W., Cheung, S., & Chen, M. (2005). Hiding privacy information in video surveillance system. In *IEEE international conference on image processing, 2005. ICIP 2005* (Vol. 3, pp. II–868–71).

Zhang, C., Rui, Y., & wei He, L. (2006). Light weight background blurring for video conferencing applications. In *2006 IEEE international conference on image processing* (pp. 481–484).

Zhang, C., Tian, Y., & Capezuti, E. (2012). Privacy preserving automatic fall detection for elderly using RGBD cameras. In K. Miesenberger, A. Karshmer, P. Penaz, & W. Zagler (Eds.), *Computers helping people with special needs. Lecture notes in computer science* (Vol. 7382, pp. 625–633). Berlin, Heidelberg: Springer.

Zhao, Q. A., & Stasko, J. T. (1998). The awareness-privacy tradeoff in video supported informal awareness: A study of image-filtering based techniques. Technical Report Georgia Institute of Technology. GVU Technical Report; GIT-GVU-98-16.

Zivkovic, Z. (2004). Improved adaptive gaussian mixture model for background subtraction. In *Proceedings of the 17th international conference on pattern recognition, 2004. ICPR 2004* (Vol. 2, pp. 28–31).