Image super-resolution for outdoor digital forensics. Usability and legal aspects

Salvador Villena[a], Miguel Vega[a], Javier Mateos[b,*], Duska Rosenberg[c], Fionn Murtagh[d], Rafael Molina[b], Aggelos K. Katsaggelos[e]

[a]*Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Granada, Spain*
[b]*Dpto. de Ciencias de la Computación e I. A., Universidad de Granada, Spain*
[c]*iCOM Research, Richmond, UK*
[d]*Centre for Mathematics and Data Science, University of Huddersfield, UK*
[e]*Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL60208-3118, USA*

Abstract

Digital Forensics encompasses the recovery and investigation of data, images, and recordings found in digital devices in order to provide evidence in the court of law. This paper is devoted to the assessment of digital evidence which requires not only an understanding of the scientific technique that leads to improved quality of surveillance video recordings, but also of the legal principles behind it. Emphasis is given on the special treatment of image processing in terms of its handling and explanation that would be acceptable in a court of law. In this context, we propose a variational Bayesian approach to multiple-image super-resolution based on Super-Gaussian prior models that automatically enhances the quality of outdoor video recordings and estimates all the model parameters while preserving the authenticity, credibility and reliability of video data as digital evidence. The proposed methodology is validated both quantitatively and visually on synthetic videos generated from single images and real-life videos and applied to a real-life case of damages and stealing in a private property.

Keywords: super-resolution, outdoor surveillance, usability, legal aspects

*Corresponding author

Email addresses: svillena@ugr.es (Salvador Villena), mvega@ugr.es (Miguel Vega), jmd@decsai.ugr.es (Javier Mateos), research@icomict.org (Duska Rosenberg), f.murtagh@hud.ac.uk (Fionn Murtagh), rms@decsai.ugr.es (Rafael Molina), aggk@eecs.northwestern.edu (Aggelos K. Katsaggelos)

1. Introduction

Digital Forensics encompasses the recovery and investigation of data, images and recordings found in digital devices in order to provide evidence in the court of law (cf. [1–6]). Digital evidence can be obtained from any devices capable of storing digital data, but there are strict national and international guidelines for its use in criminal and civil investigations as part of legal processes. The most common are the British ACPO [7] and US NIJ guidelines for the appropriate use of digital evidence [8] that includes gathering digital data, processing of digital data and the preparation of digital data to be presented in courts by both forensic and legal professionals.

Most of the Digital Forensics literature focuses on practical activities of gathering digital evidence, preparing it for presentation in courts and presenting it in court by legal professionals and expert witnesses. Issues of authenticity, reliability and credibility addressing the concerns of the legal professions have been raised and operational standards and structured processes devised in order to resolve them. They have provided regulation in digital forensic practice, but more needs to be done.

As well as being used to prove that a criminal act has been committed, digital evidence is required to aid in identification of the perpetrators, confirming alibis, identifying sources of documents and confirming their authenticity. There are, however, significant issues still to be addressed related to not only the increasing size of digital media, but also the complexity of their use. These complexities arise from an increasing number of users owning multiple devices capable of storing and sharing potential digital evidence. In addition, the integration of digital evidence (and scientific evidence in general) is often influenced by social, cultural and religious factors that underpin legal systems in different countries (cf. [9]).

To generate those High Resolution (HR) images of a scene, a general variational Bayesian approach to the super-resolution (SR) problem [10–12] is proposed. For the first time, the general modelling of Super Gaussian (SG) distributions [13] is applied to SR. SGs are priors capable of capturing the sparse distribution of edges within natural images. SG priors have been successfully applied to blind image deconvolution [14–16]. Here, they are combined with proper modelling of the observation process as well as the registration parameters in order to obtain a high quality HR image from a set of LR observations.

The rest of the paper is divided as follows. In the next section, a bibliographic study on the status of digital evidence and SR is presented. Section 3 introduces the SR problem and formulates it using the Bayesian framework. Then, in Section 4 a solution to the SR problem using variational Bayesian inference is proposed. The proposed methodology is applied, in Section 5, to the study of a real-life forensic case to help identify the culprits of damages in a property and synthetic video sequences generated from an image and a real-life video which allow to compare the resulting images both visually and numerically. Finally, Section 6 concludes the article.

2. State-of-the-Art

Many studies of digital evidence in legal practice are focused on the reliability and acceptability of digital evidence as shown by categorisations of Levels of Certainty that were

devised by Casey [3]. Therefore, in order to discover, examine and provide evidence to legal enforcement in criminal events, a wide range of issues need to be addressed [4]. As digital forensics is essentially a process of applying scientific methods to the discovery, examination and provision of evidence to legal enforcement in criminal events [17], the credibility of digital evidence requires not only understanding of the scientific techniques but also understanding of the legal principles. The procedures that traditionally safeguard the integrity of evidence, including digital evidence, involve establishing that an incident has occurred, determining the nature of the incident and identifying the culprits (as far as this is possible under specific circumstances). Unlike physical evidence, digital evidence is often regarded by legal professionals as fragile, that is, it can be lost, altered, damaged, or accessed by unauthorised personnel. It is therefore of critical importance that forensic investigations safeguard its integrity by exercising evidential controls, such as maintaining the chain of custody as well as ensuring that it is gathered and protected through structured processes that are acceptable to the courts. "Tainted evidence that may have been acquired or protected without the requisite level of security may be legally inadmissible." [18]. Guidance on the process of analysing and interpreting digital evidence is also necessary as it provides the structure to the analytical and interpretational processes so that different investigators working on the same digital evidence can obtain the same results. Furthermore, any changes to the digital evidence in the process of analysis and interpretation should be traceable and justifiable in order to preserve the credibility of both the evidence and the analyst in the eyes of legal professionals. This is quite a challenge given the volume, variety and complexity of digital evidence, and raises issues of selection and use of forensic tools as well as proficiency and competency of the investigators themselves [19].

In this context, video as evidence has to be authenticated so that it is clear whether it is original or an altered copy since the nature of the alteration may render it inadmissible in court. This could happen if, for example, it cannot be proved that in spite of the alterations the video still depicts the scene of the crime and that the location, date and time when the recording was taken have remained the same as in the original. Traditional approaches of evidential control, described briefly above, may not be sufficient to guarantee the authenticity of the video as evidence [20]. However, digital systems usually provide methods for authentication such as metadata or serial numbers hidden in the video [21] as well of stronger forms of identification based on the image or video itself. Sensors imperfections and noise, photo response non-uniformity [22] or defective pixels can help to authenticate digital images and videos (see [23, 24]).

Nowadays, surveillance cameras are ubiquitous and their recorded videos are often used to identify the perpetrator. However, surveillance cameras usually suffer from poor quality and low resolution which prevent identification on the frames as extracted from the recorder. Image SR can help bridge the gap between poor video quality and evidence gathering [25]. The image SR problem has received a lot of attention from the image processing and computer vision research community in the past two decades (see [26–29] for a review). We can distinguish between *Multiple-Image Super-resolution (MISR)* and *Single Image Super-resolution (SISR)*. SISR [30] allows to obtain a HR image from only one observed LR image by applying, for instance, interpolation [31–34] or machine learning techniques based on
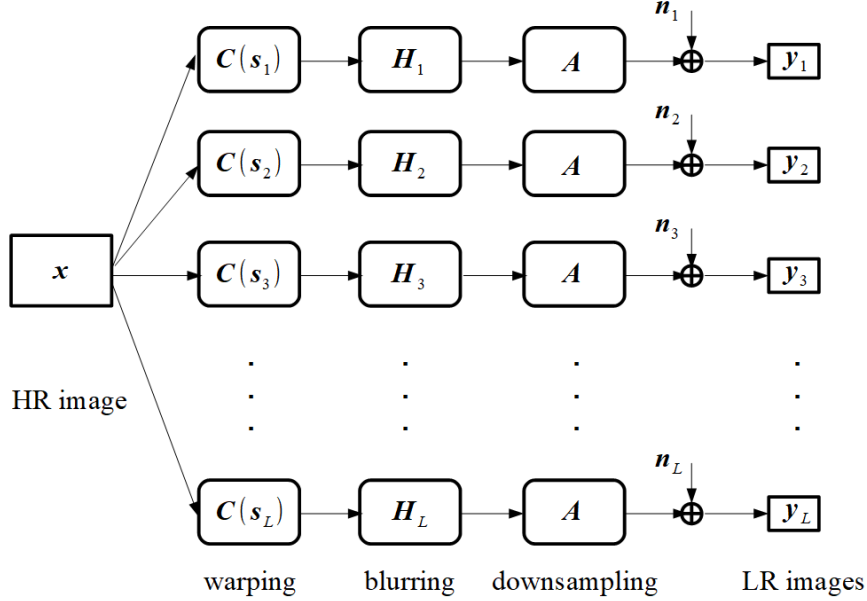
Figure 1: Illustration of the degradation model in Eq. (1).

LR/HR image patches [35–39], see [40] for the use of deep learning techniques in image recovery problems. However, when a video sequence or a set of LR images is available, MISR is preferred.

MISR allows to infer a spatially HR image of a scene, from multiple LR images affected by warping, blurring, and the noise inherent to the capture process [41]. Frames of a video sequence may contain many small shifted or rotated LR images of a given object, caused by the acquisition process, and the camera and/or scene motion, from which a HR image can be obtained using MISR techniques. MISR can be applied to obtain either a single HR image from a sequence of many LR images or a HR image sequence from a LR image sequence [42–45].

Although some SR algorithms with application to forensic investigation [46–51] have been proposed (see [46–51]), they are mainly formulated from an image processing point-of-view. A few briefly discuss on the use in a court of law of SR (see [25] for instance) or image enhancement techniques in general [52, 53] but, to the best of our knowledge, no-one discusses in depth the forensic aspects or analyses real-life cases where SR had an important role to play.

3. Problem Formulation

Let us now describe the MISR problem, i.e., the reconstruction of a HR image $\mathbf{x}$ from a sequence of L LR observed images $\mathbf{y} = \{\mathbf{y}_k\}$, $k = 1, \ldots, L$, of the same scene.

Each LR image $\mathbf{y}_k$ consists of $N = N_h \times N_v$ pixels while the size of the HR image $\mathbf{x}$ is PN, where $\sqrt{P} \in \mathbb{N}$ is the factor of increase in resolution. In this paper we adopt the matrix-vector notation, images $\mathbf{y}_k$ and $\mathbf{x}$ are arranged as $N \times 1$ and $PN \times 1$ column vectors, respectively. The imaging process, illustrated in Fig. 1, introduces warping, blurring and downsampling, which is modelled as

$$\mathbf{y}_k = \mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)\mathbf{x} + \mathbf{n}_k = \mathbf{B}_k(\mathbf{s}_k)\mathbf{x} + \mathbf{n}_k, \tag{1}$$

where $\mathbf{A}$ is the $N \times PN$ downsampling matrix, $\mathbf{H}_k$ is the $PN \times PN$ matrix modelling sensor integration and blurring, $\mathbf{C}(\mathbf{s}_k)$ is the $PN \times PN$ warping matrix generated by the motion vector $\mathbf{s}_k$, and $\mathbf{n}_k$ is the $N \times 1$ acquisition noise. A detailed description of the explicit form of the warping matrices $\mathbf{C}(\mathbf{s}_k)$ in Eq. (1) can be found in [11]. The effects of downsampling, blurring, and warping are combined into the $N \times PN$ system matrix $\mathbf{B}_k(\mathbf{s}_k) = \mathbf{A}\mathbf{H}_k\mathbf{C}(\mathbf{s}_k)$, from which each row maps the pixels of the HR image $\mathbf{x}$ to a given pixel in the LR image $\mathbf{y}_k$.

Given Eq. (1), the SR problem is expressed as the search of an estimate of the HR image $\mathbf{x}$ from the set of LR images $\mathbf{y}$ using our prior knowledge on $\{\mathbf{C}(\mathbf{s}_k)\}$, $\{\mathbf{n}_k\}$, and $\mathbf{x}$.

The ill-posed nature of the SR problem has been traditionally circumvented by means of regularisation terms in the optimisation approach, or using prior distributions in the Bayesian approach (see [26]).

3.1. Observation Model

Using the model in Eq. (1) and assuming that $\mathbf{n}_k$ is zero-mean white Gaussian noise with inverse variance (precision) β_k, the conditional distribution of each LR image $\mathbf{y}_k$ is given by

$$\mathrm{p}(\mathbf{y}_k|\mathbf{x}, \mathbf{s}_k, \beta_k) \propto \beta_k^{N/2} \exp\left[-\frac{\beta_k}{2} \parallel \mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x} \parallel^2\right]. \tag{2}$$

Assuming statistical independence of the noise among the LR image acquisitions, as already assumed in most of the existing super-resolution methods [54–58], the conditional probability of the set of LR images $\mathbf{y}$ given the unknown HR image $\mathbf{x}$ and the model parameters can be expressed as

$$\mathrm{p}(\mathbf{y}|\mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\}) = \prod_{k=1}^{L} \mathrm{p}(\mathbf{y}_k|\mathbf{x}, \mathbf{s}_k, \beta_k). \tag{3}$$

3.2. Image Prior Model

In this paper we will use SG distributions [13] as priors. Those distributions have the form

$$\mathrm{p}(s) = Z \exp\{-\rho(s)\}, \tag{4}$$

with $\rho(s) : \mathbb{R} \to [-\infty, \infty]$ a penalty function symmetric around zero and $\rho(\sqrt{s})$ concave and increasing for $s \in [0, \infty)$. This condition is equivalent to $\rho'(s)/s$ being decreasing on $(0, \infty)$.

Table 1: Different choices for the penalty function $\rho(s)$.

Label	$\rho(s)$	$\rho'(s)/|s|$
ℓ_p , $0 < p \leq 2$	$\frac{1}{p}|s|^p$	$|s|^{p-2}$
log	$\log(\epsilon + |s|)$	$(\epsilon + |s|)^{-1}|s|^{-1}$

Table 1 shows some examples of energy functions $\rho(\cdot)$ associated with SG distributions. SG distributions promote sparsity [14]. The energy, $\rho(\cdot)$, associated with an SG distribution can be represented as (see [59])

$$\rho(s) = \inf_{\xi > 0} \frac{1}{2}\xi s^2 - \rho^*(\frac{\xi}{2}) \,, \tag{5}$$

where $\rho^*(\cdot)$ is the concave conjugate function of $\rho(\cdot)$. Eq. (5) provides a quadratic upper bound to the energy of an SG distribution which naturally leads to a Gaussian approximation.

Having defined the SG distributions in general, let us use them in the SR problem at hand. Let us define

$$\mathbf{z}_j = \mathbf{F}_j\mathbf{x} \quad \text{for} \quad j = 1,..,d \,. \tag{6}$$

where the $\mathbf{F}_j$ are convolution operators, we use first and second order differences, and d is the number of filters used, and define $\{\mathbf{z}\} = \{\mathbf{z}_1,..,\mathbf{z}_d\}$. We propose the following prior distribution over the unknown $\mathbf{x}$

$$\mathrm{p}(\mathbf{x}) \propto \prod_{j=1}^{d}\prod_{i=1}^{PN} \mathrm{p}(z_j(i)) \,, \tag{7}$$

where $\mathrm{p}(z_j(i))$ are SGs as defined in (4). Using (5) the following lower bound to $\mathrm{p}(\mathbf{x})$ can be obtained

$$\mathrm{p}(\mathbf{x}) \geq \mathrm{p}(\mathbf{x}|\boldsymbol{\xi}) = Z\prod_{j=1}^{d}\prod_{i=1}^{PN} \exp\left[-\left\{\frac{1}{2}\xi_j(i)\parallel z_j(i)\parallel^2 - \rho^*(\frac{\xi_j(i)}{2})\right\}\right] \tag{8}$$

where the components of the variational parameter vector $\boldsymbol{\xi} = \{\boldsymbol{\xi}_1,..,\boldsymbol{\xi}_d\}$ have already been introduced.

3.3. Modelling the registration parameters

In this paper the uncertainties on the registration parameters $\{\mathbf{s}_k\}$ are modelled following [11]. We denote by $\{\bar{\mathbf{s}}^p\}$ the estimate of $\mathbf{s} = \{\mathbf{s}_1,\ldots,\mathbf{s}_L\}$ obtained from the LR observations in a preprocessing step, using registration algorithms, see, for instance, [60]. As these estimates are in general inaccurate, we model the motion parameters as stochastic variables

following Gaussian distributions with *a priori* means set equal to the preliminary motion parameters $\bar{\mathbf{s}}_k^p$, $k = 1, \ldots, L$, that is,

$$p(\{\mathbf{s}_k\}) = \prod_{k=1}^{L} \mathcal{N}(\mathbf{s}_k | \bar{\mathbf{s}}_k^p, \mathbf{\Xi}_k^p), \tag{9}$$

with $\mathbf{\Xi}_k^p$ the *a priori* covariance matrix. The parameters $\bar{\mathbf{s}}_k^p$ and $\mathbf{\Xi}_k^p$ incorporate prior knowledge about the motion parameters into the estimation procedure. If such knowledge is not available, $\bar{\mathbf{s}}_k^p$ and $(\mathbf{\Xi}_k^p)^{-1}$ can be set equal to zero, which makes the observations solely responsible for the estimation process.

3.4. Hyperpriors on the Hyperparameters

In this paper we assume flat hyperpriors for the $\xi_j(i)$ hyperparameters in Eq. (8). That is $p(\xi_j(i)) \propto const$. To model the hyperparameters in Eq. (3), we employ Gamma distributions

$$p(\{\beta_k\}) = \prod_{k=1}^{L} \Gamma(\beta_k | a_{\beta_k}^o, b_{\beta_k}^o), \tag{10}$$

where $a_{\beta_k}^o > 0$ and $b_{\beta_k}^o > 0$ are the shape and scale parameters, respectively. The hyperpriors are chosen as Gamma distributions since they are conjugate priors for the Gaussian distribution.

3.5. Joint Model

Combining Eqs. (3), (7), (9) and (10) we obtain the following lower bound to the joint probability distribution

$$\begin{aligned} p(\Theta, \mathbf{y}) &\geq p(\Theta, \boldsymbol{\xi}, \mathbf{y}) \\ &= p(\mathbf{y} | \mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\}) p(\mathbf{x} | \boldsymbol{\xi}) p(\{\beta_k\}) p(\{\mathbf{s}_k\}), \end{aligned} \tag{11}$$

where $\Theta = \{\mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\}\}$.

4. Bayesian Inference

The Bayesian inference is based on the posterior distribution $p(\Theta \mid \mathbf{y}) = \frac{p(\Theta, \mathbf{y})}{p(\mathbf{y})}$. As $p(\mathbf{y})$ cannot be obtained, we approximate $p(\Theta \mid \mathbf{y})$ by the distribution, $q(\Theta)$, having the minimum Kullback-Leibler divergence to $p(\Theta \mid \mathbf{y})$. This is the well known variational approximation, which is described in [61, 62] (see also [63], [57] and [10]). Within the mean field approximation, $q(\Theta)$ is assumed to have the form $q(\Theta) = \prod_{\zeta \in \Theta} q(\zeta)$, with

$$q(\zeta) \propto \exp\left(\langle \log p(\Theta, \mathbf{y}) \rangle_{\Theta_\zeta} \right), \tag{12}$$

where Θ_ζ denotes the set Θ with ζ removed, and $\mathbf{E}_{q(\Theta_\zeta)}[\cdot] = \langle \cdot \rangle_{\Theta_\zeta}$. In the following, the subscript of the expected value will be removed when it is clear from the context.
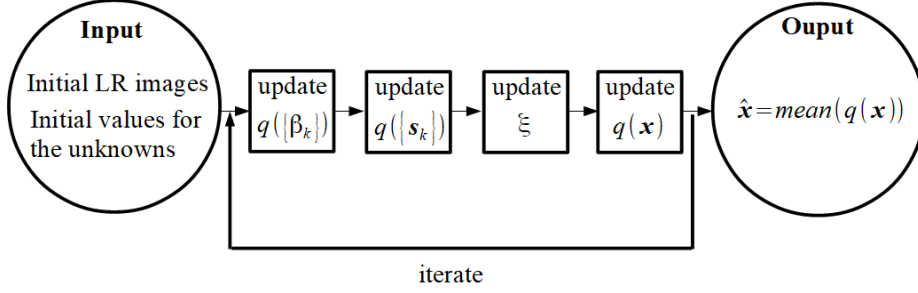
Figure 2: Graphical description of the inference process.

Instead of using $p(\Theta, \mathbf{y})$ we utilise its lower bound $p(\Theta, \boldsymbol{\xi}, \mathbf{y})$ in (11), which includes the variational parameter $\boldsymbol{\xi}$, and solve iteratively

$$\log q(\zeta) \propto \langle \log p(\Theta, \hat{\boldsymbol{\xi}}, \mathbf{y}) \rangle_{\Theta \setminus \{\theta\}} \quad \forall \zeta \in \Theta, \tag{13}$$

$$\hat{\boldsymbol{\xi}} = \arg \max_{\xi} \langle \log p(\Theta, \mathbf{y}, \boldsymbol{\xi}) \rangle_{q(\Theta)} \tag{14}$$

This procedure, depicted graphically in Fig. 2, starts each iteration by fixing the current estimates of the distribution of the unknowns $q(\mathbf{x})$ and $q(\{\mathbf{s}_k\})$ and the value of $\boldsymbol{\xi}$ and then estimates a new distribution of the hyperparameters $q(\{\beta_k\})$. This updated distribution will be used, with $q(\mathbf{x})$ and $\boldsymbol{\xi}$, to obtain a new estimate for the distribution of the registration parameters, $q(\{\mathbf{s}_k\})$, which is then used to update the variational parameters $\boldsymbol{\xi}$. Finally, the distribution of the super-resolved image $q(\mathbf{x})$ is estimated from the updated values of the rest of unknowns. This procedure is repeated iteratively until convergence.

Let us now detail the equations for solving on each unknown. From Eq. (13), we obtain for $q(\mathbf{x})$

$$q(\mathbf{x}) \propto \exp \left\{ \langle \log(p(\mathbf{y}|\mathbf{x}, \{\mathbf{s}_k\}, \{\beta_k\})) \rangle_{\{\mathbf{s}_k, \beta_k\}} + \log \left(p(\mathbf{x}|\hat{\boldsymbol{\xi}}) \right) \right\}, \tag{15}$$

which is the multivariate Gaussian

$$q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}, \boldsymbol{\Xi}_{\mathbf{x}}), \tag{16}$$

with

$$\boldsymbol{\Xi}_{\mathbf{x}}^{-1} = \sum_{j=1}^{d} \mathbf{F}_j^t \operatorname{diag}(\hat{\boldsymbol{\xi}}_j) \mathbf{F}_j + \sum_{k=1}^{L} \langle \beta_k \rangle \langle \mathbf{B}(\mathbf{s}_k)^t \mathbf{B}(\mathbf{s}_k) \rangle_{\mathbf{s}_k}, \tag{17}$$

and

$$\hat{\mathbf{x}} = \boldsymbol{\Xi}_{\mathbf{x}} \sum_{k=1}^{L} \langle \beta_k \rangle \langle \mathbf{B}(\mathbf{s}_k) \rangle_{\mathbf{s}_k}^t \mathbf{y}_k, \tag{18}$$

8

which can be solved iteratively utilising a *Conjugate Gradient* method.

Also, from Eq. (12), we find the following distribution for the registration parameters

$$q(\mathbf{s}_k) \propto \exp\left[-\frac{1}{2}\left(\langle\beta_k\rangle\langle\|\ \mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\ \|^2\rangle_{\mathbf{x}}\right.\right.$$
$$\left.\left.+\ (\mathbf{s}_k - \bar{\mathbf{s}}_k^p)^t\,(\mathbf{\Xi}_k^p)^{-1}\,(\mathbf{s}_k - \bar{\mathbf{s}}_k^p)\right)\right]\ . \tag{19}$$

The explicit form of the distribution $q(\mathbf{x})$ in Eq. (16) depends on the expectation values $\langle\mathbf{B}(\mathbf{s}_k)^t\mathbf{B}(\mathbf{s}_k)\rangle_{\mathbf{s}_k}$ and $\langle\mathbf{B}(\mathbf{s}_k)\rangle_{\mathbf{s}_k}$, and $q(\mathbf{s}_k)$ in Eq. (19) depends on $\langle\|\ \mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\ \|^2\rangle_{\mathbf{x}}$. These calculations are not straightforward since $\mathbf{C}(\mathbf{s}_k)$, in Eq. (1), is nonlinear with respect to $\mathbf{s}_k$. Therefore, we expand $\mathbf{C}(\mathbf{s}_k)$ using its first-order Taylor series around the mean value $\bar{\mathbf{s}}_k = \langle\mathbf{s}_k\rangle = (\bar{\theta}_k, \bar{c}_k, \bar{d}_k)^T$ of the distribution $q(\mathbf{s}_k)$, in Eq. (19) (details can be found in [11]).

4.1. Estimation of the Variational Parameter $\boldsymbol{\xi}$

Using (14), we obtain for the variational parameters

$$\hat{\xi}_j(i) = \arg\min_{\xi_j(i)} \frac{1}{2}\xi_j(i)\langle z_j^2(i)\rangle_{\mathbf{x}} - \rho^*\left(\frac{\xi_j(i)}{2}\right) \tag{20}$$

with

$$\langle z_j^2(i)\rangle_{\mathbf{x}} = \hat{\mathbf{x}}^t\mathbf{F}_j^t\mathbf{J}^{ii}\mathbf{F}_j\hat{\mathbf{x}} + \mathrm{Tr}(\mathbf{\Xi}_{\mathbf{x}}\mathbf{F}_j^t\mathbf{J}^{ii}\mathbf{F}_j)\,, \tag{21}$$

where $\mathbf{J}^{ii}$ is a single-entry $PN \times PN$ matrix with zeros everywhere except at the entry (i,i), which is equal to one. In this paper, $\mathbf{\Xi}_{\mathbf{x}}$ in Eq. (21) is calculated by applying the Jacobi approximation.

The minimum in (20) is achieved at (see [14])

$$\hat{\xi}_j(i) = \rho'\left(\sqrt{\langle z_j^2(i)\rangle_{\mathbf{x}}}\right)/\sqrt{\langle z_j^2(i)\rangle_{\mathbf{x}}}\,. \tag{22}$$

4.2. Estimation of the hyperparameter distributions

Finally we obtain the distributions for the hyperparameters $\{\beta_k\}$, which are found to be Gamma distributions. For the $\{\beta_k\}$ hyperparameters, using Eq. (13), we obtain

$$q(\beta_k) \propto \beta_k^{\frac{N}{2}-1+a_{\beta_k}^0} \exp\left[-\beta_k\left(b_{\beta_k}^0 + \frac{\langle\|\mathbf{y}_k - \mathbf{B}_k(\mathbf{s}_k)\mathbf{x}\|^2\rangle_{\mathbf{x},\mathbf{s}_k}}{2}\right)\right] \tag{23}$$

with

$$\langle\beta_k\rangle = \frac{N + 2a_{\beta_k}^0}{\|\mathbf{y}_k\mathbf{B}(\mathbf{s}_k)\hat{\mathbf{x}}\|^2 + \mathrm{Tr}(\mathbf{\Xi}_{\mathbf{x}}\mathbf{B}^t(\mathbf{s}_k)\mathbf{B}(\mathbf{s}_k)) + 2b_{\beta_k}^0}\,. \tag{24}$$

Algorithm 1 summarises the proposed iterative SR method which comprises the HR image estimation, as well as the estimation of the registration parameters, the HR prior parameters and the observation model hyperparameters.

9

Algorithm 1 Variational Bayesian Super-resolution

Require: : Low resolution image set $\{\mathbf{y}_k\}$, values $\{a_{\beta_k}^0\}$, $\{b_{\beta_k}^0\}$, $\{\bar{\mathbf{s}}_k^p\}$, $\{\boldsymbol{\Xi}_k^p\}$ and initial HR image estimation $\hat{\mathbf{x}}^{(0)}$.

Set $n = 0$, $q^{(0)}(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}^{(0)}, \mathbf{0})$ and $q^{(0)}(\{\mathbf{s}_k\}) = p(\{\mathbf{s}_k\})$, with $p(\{\mathbf{s}_k\})$ given in Eq. (9).

while convergence criterion is not met **do**

    0. Set $n = n + 1$

    1. Given $q^{(n-1)}(\mathbf{x})$ and $q^{(n-1)}(\{\mathbf{s}_k\})$, obtain $q^{(n)}(\{\beta_k\})$ using Eq. (23).

    2. Given $q^{(n-1)}(\mathbf{x})$ and $q^{(n)}(\{\beta_k\})$, obtain $q^{(n)}(\{\mathbf{s}_k\})$ using Eq. (19).

    3. Given $q^{(n-1)}(\mathbf{x})$, compute $\hat{\boldsymbol{\xi}}^{(n)}$ using Eq. (22).

    4. Given $q^{(n)}(\{\beta_k\})$, $q^{(n)}(\{\mathbf{s}_k\})$ and $\hat{\boldsymbol{\xi}}^{(n)}$, obtain $q^{(n)}(\mathbf{x})$ using Eq. (16).

end while

Output the HR image $\hat{\mathbf{x}}$ as the mean of $q^{(n)}(\mathbf{x})$, see Eq. (18).

---

5. Experimental validation

Experimental results are presented on synthetic video sequences and a real-life video sequence of a case of damages and stealing in a private property. The synthetic experiments allow to quantitatively evaluate the quality of the obtained images by comparing them to the real underlying ones. For the real-life video sequence, we first describe how image processing helps in providing digital evidences and the steps required to bring digital evidence into a court and, then, show how we proceeded to obtain a super-resolved image of the person of interest in the video.

5.1. Validation on Synthetic Sequences

While tests on real video sequences will later confirm the applicability of the proposed methodology to real-life cases, it is not possible to provide metric analysis of the processed images. In order to provide a quantitative evaluation, the proposed methodology has been tested on synthetic video sequences simulating the kind of images taken by surveillance cameras. Two different set of tests have been performed. In the first one, a sequence is synthesised from a single image by translating, rotating and subsampling it. The second one simulates a low resolution video by downsampling a high resolution video sequence under conditions similar to the ones found in regular surveillance cameras.

First, we tested the proposed approach on a set of low resolution images generated from a single high resolution image that serves as reference frame. Results are reported on the two different standard test images shown in Fig. 3. A set of five degraded low resolution images were generated from the original images following the degradation model in Eq. (1) with rotations between ± 5 degrees and subpixel translations, a 3×3 uniform blur to simulate sensor integration and a downsampling by a factor of two by discarding every other pixel. Zero mean Gaussian noise was added to obtain different SNRs ranging from 10 to 40dB, thus obtaining images simulating different scenarios from a low luminosity/night scene where noise is dominant due to high camera gain to a well illuminated one. The images, of which
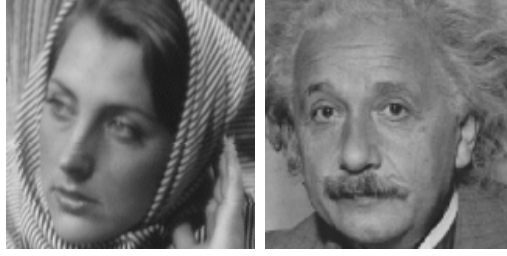
Figure 3: Original Barbara and Einstein test images.

an upsampled example is shown in the first row of Figs. 4 and 5, were input to the SR methods to obtain estimates of the original high resolution images.

In order to assess the quality of the proposed methodology, we compare the outcome of bicubic interpolation, the method in [64] and the proposed methods. Resulting images, for visual inspection are shown in Figs. 4 and 5 while quantitative results, by means of the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) [65] measures, are presented in Fig. 6 for an easy comparison. Bicubic interpolation is a basic technique which is usually used when the information is already present in a single frame but its size is too small for a correct visualisation and hence zooming is needed. Interpolated images, depicted in the second row of Figs. 4 and 5, are smoothed versions of the observations and obtain the lowest PSNR and SSIM values for both images. The method proposed by Molina *et al.* [64], which utilizes a simultaneous autoregressive (SAR) prior model for the image, resulted in the images displayed in the third row of Figs. 4 and 5. This model is a particular case of super-Gaussian priors where the ℓ_2 penalty function is used and only one convolution filter corresponding to the Laplacian operator is considered in Eq. (6). Resulting images show a higher level of detail than bicubic interpolation but, also, piecewise smoothing and artifacts, specially at low SNR, which spoil its quality and the PSNR and SSIM figures-of-merit.

The rest of the results in Figs. 4 and 5 were obtained using the proposed super-resolution algorithm, summarized in Alg. 1. The initial parameters were set as follows: The values of $\{a^0_{\beta_k}\}$ and $\{b^0_{\beta_k}\}$, $k = 1, \ldots, 4$, which encode the prior information about the noise parameters β_k were set to a value very close to zero in Eq. (10). With those values, the variance of the Gamma distributions, $\{a^0_{\beta_k}/(b^0_{\beta_k})^2\}$, are huge, letting the method to automatically estimate the noise parameter from the observations. The prior registration parameters $\{\bar{\mathbf{s}}^p_k\}$ and the inverse of the *a priori* covariance matrices, $(\boldsymbol{\Xi}^p_k)^{-1}$, were set to zero, hence making the data fully responsible of the estimation of the registration parameters. Bicubic interpolation was used as the initial high resolution image $\hat{\mathbf{x}}^{(0)}$. The algorithm was run until the convergence criterion $\| \hat{\mathbf{x}}^{(i)} - \hat{\mathbf{x}}^{(i-1)} \|^2 / \| \hat{\mathbf{x}}^{(i)} \|^2 < 10^{-5}$ is met. The integration and blurring matrix $\mathbf{H}_k$ was set to the convolution matrix obtained from a 3×3 uniform kernel. For the prior distribution on the high resolution image in Eq. (7), we used the horizontal, vertical, upper-right diagonal and lower-right diagonal first order difference convolution operators. We want to note that the method is both fully automatic, not needing of human intervention for parameter tuning, and very flexible, allowing to easily incorporate prior knowledge about the noise or the registration parameters if available.
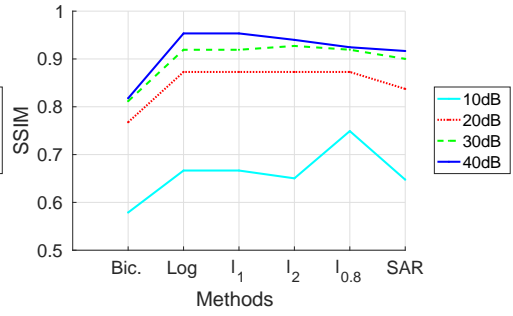
11

Figure 4: Super-resolved Barbara image using different methods and noise levels.
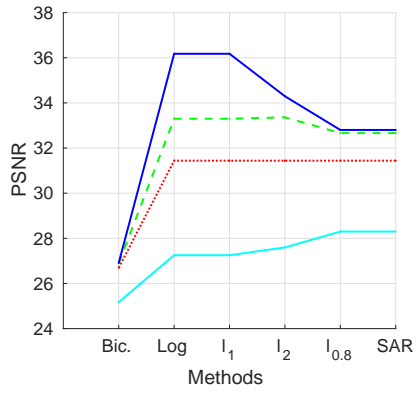
Figure 5: Super-resolved Einstein image using different methods and noise levels.
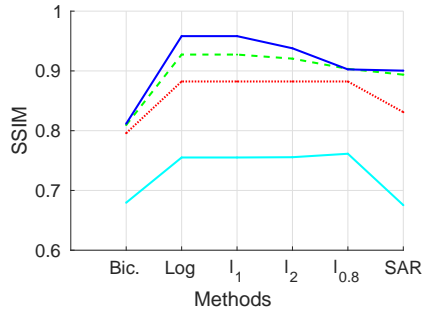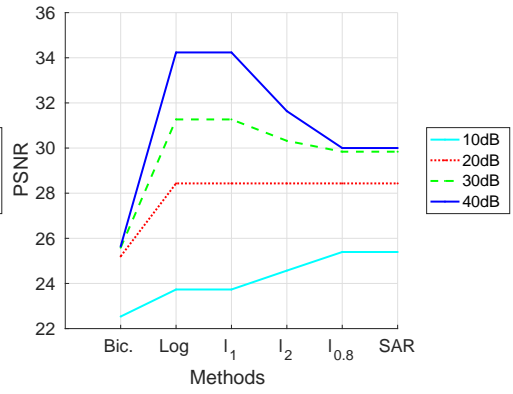
(a) SSIM for Einstein   (b) SSIM for Barbara

(c) PSNR for Einstein   (d) PSNR for Barbara

Figure 6: PSNR and SSIM comparison of the different SR methods for the Barbara and Einstein images with different noise levels.

Figure 7: A frame of the synthetic video with the detected face marked.

| (a) HR frame | (b) LR frame | (c) Bicubic interp. | (d) Method in [64] |

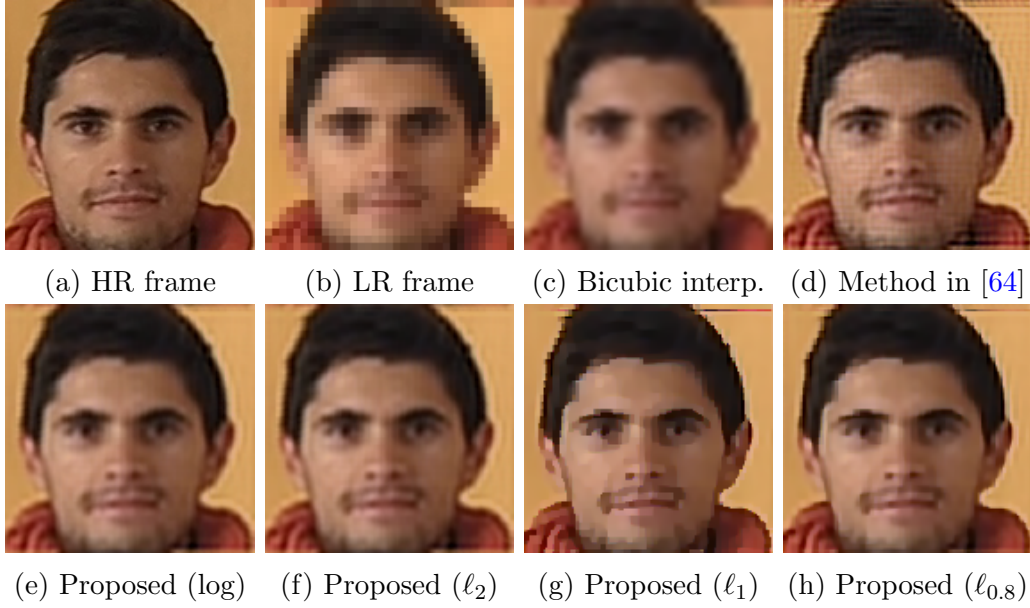| (e) Proposed (log) | (f) Proposed (ℓ_2) | (g) Proposed (ℓ_1) | (h) Proposed ($\ell_{0.8}$) |

Figure 8: An observed frame of the synthetic video sequence and its reconstruction using different methods.

Results from the proposed methods, are visually and numerically much more accurate than those provided by bicubic interpolation and the SAR method in [64], with an important increase of both PSNR and SSIM values (see Fig. 6). Results with the log and ℓ_1 priors are almost identical so only log results are shown in Figs. 4 and 5. Depending on the amount of noise, one of the proposed methods may be preferable over the others. If noise is high, the $\ell_{0.8}$ method (see the first column of the last row of Figs. 4 and 5) provides better results since it assumes that the image is piecewise flat so it removes the noise while preserving the edges of the images. However, as the noise decreases, the $\ell_{0.8}$ prior also removes fine details, hence producing not very pleasant reconstructions (see, for instance, that the hair and face of Einstein in Fig. 5 appear oversmoothed). The proposed method using the log and ℓ_1 penalty functions, on the other hand, are able to recover those details, but cannot handle a large amount of noise in the images. When a moderate amount of noise is present on the image, all the proposed methods perform similarly and better than bicubic and SAR methods.

15

	Bicubic interp.	method in [64]	proposed (log)	proposed (ℓ_2)	proposed (ℓ_1)	proposed ($\ell_{0.8}$)
PSNR	25.71	27.99	28.63	28.63	25.71	27.55
SSIM	0.777	0.843	0.872	0.872	0.857	0.881

Table 2: PSNR and SSIM figures-of-merit for the synthetic video sequence

Additionally, we tested the proposed methods on videos which simulate the recording of an assault. The videos were obtained using a full-HD (1920x1080 pixels) camera and degraded by a 3×3 uniform blur to simulate sensor integration and a downsampling by a factor of four in each direction. To select the input frame portions for the algorithm, faces have to be located in the video. Face detection algorithms (see [66] for a recent review applied to surveillance) can be used to help the investigators to discover the location of persons in the set of video frames. This will save a large amount of time watching the video in search of the exact fragment where faces appear. However, the investigator still needs to select the best pose and the most promising frames to feed the MISR algorithm and also interpret the results. For this synthetic experiment, we run the Viola-Jones face detection algorithm [67] to find the bounding box where faces are located. An example of a frame of the video with the detected face marked in red is shown in Fig. 7. For this simulation, we selected seven of the most promising frames from the ones with a face detected since, in our experience, good quality frames are scarce in real sequences.

The super-resolution methods were run on those frame portions to obtain the estimated HR images depicted in Fig. 8. The HR reference image and an upsampled LR image are depicted in Fig. 8a and Fig. 8b, respectively, as a reference. Again, the proposed methods provide much sharper and detailed images than bicubic interpolation and the method in [64], without noticeable artifacts. Numerical results, shown in Table 2, confirm this quality increase. The proposed methods increase the PSNR in about 3dB and the SSIM in almost a tenth, with respect to bicubic interpolation. Note that we magnified the images by factor of 4 in each direction ($P = 16$) using only 7 LR frames. For this particular sequence, adding more LR frames increases slightly the quality of the reconstruction but, in general, it might be difficult to find a significant number of frames with the person of interest in almost the same pose. Including different poses increases the probability of appearance of artifacts that could challenge SR applicability.

As a final note, we want to emphasize that the proposed pipeline allows the detection of the frames with faces and the automatic reconstruction although the investigator needs to select the most promising frames. We are studying the possibility of including outlier detection techniques (see [68], for instance) into the super-resolution algorithm to alleviate this problem. Note also, that the proposed methodology maintains the integrity of evidence as video sequences can be processed with different methods without tainting the original frames.

5.2. A Real-life Case Study of Outdoor Surveillance

Image and video processing are powerful tools in digital forensics which allow to enhance recorded input images to obtain evidence. When a criminal offence is committed, the victim needs to report it to the police, which will initiate investigation. Digital forensic experts will help the police with the investigation by providing supporting evidence and processing the digital evidence according to specific guidelines.

The empirical research method to discover and analyse responses by legal professionals to the original video recordings is based on social science investigations involving observation, interviews and documentary evidence (including video recordings). The discovery stage begins with "use-case scenario" describing the situation as closely as possible using the informant's (here, the victim's) own words. The discovery data was obtained from interviews and observations of the scene as described in Section 5.2.1. The investigation process, described in Section 5.2.2, is based on documentary evidence involving the internationally accepted guidelines from ACPO, outlining how the digital forensic investigation should be conducted. The three data sources (interviews, observations and documents) thus provided the context for the analysis of the authenticity, credibility and reliability of the video recordings as digital evidence.

5.2.1. Report to the Police of a Criminal Offence

The first step is reporting the criminal offence to the police. In the report, special effort is needed in describing the surveillance system, position of the cameras to ensure privacy of neighbouring properties and damages committed by perpetrators. In the real-life case we study, surveillance was performed by cameras (one visible and one hidden) recording the outside area of the victim's home. Report to the police includes damages to the fence committed by perpetrators identified by the victim. The victim, who describes the process as follows, looked for help of image processing researchers to perform digital investigation.

"My holiday home has been damaged several times over the past couple of years, but the perpetrators have never been identified. I put surveillance cameras and told the local police that I had two cameras taking videos of the terrace and the fence that were under continuous attack. I put one camera in a visible place on the terrace (a dummy bought at Amazon for £10.00) one real camera behind the glass window, which could be seen if you looked carefully and the second real camera was hidden inside the porch roof just above the outdoor lamp so that the two together looked like one. Nobody knew about the second real camera. The legal constraints were that the cameras should be positioned so that they record my space but do not intrude into the neighbour's private space, and that there were visible signs saying that the house was under surveillance.

The first recording was taken in the middle of the night, looked really spooky, showing white sticks moving in front of the camera, as well as demonstrating the quality of the images in night time (see Fig. 9). It turned out to be a spider weaving his web. The second recording was of a hooded man with a scarf covering his face, all camouflaged, as shown in Fig. 10(a). He covered the dummy camera with a bucket, and used a stick to close the shutters to cover the first real camera. Then he stood in front of the second (hidden) real camera (Fig. 10(b)) and took his hood and scarf off. The rest of the recordings showed

17

Figure 9: Two examples of night vision camera. White areas corresponds to a spider weaving his web.

three young men breaking down the fence and stealing the garden gate as can be seen in Figs. 10(c) and 10(d).

I reported it to the police who did nothing for months. Since it was clear from the recordings that criminal damage was being done, the perpetrators do not have the right for privacy protection and I am free to show the recordings in public."

5.2.2. Digital Forensic Investigation

The police are then expected to carry out an investigation in order to identify the culprits. There are also digital forensic experts to advise how digital evidence can be strengthened with traditional empirical methods, such as going to suspects' homes to find clothes they were wearing, or for example, finding photographs of the perpetrators on the web pages wearing the same clothes. Other digital methods can also be used, for example, measuring their height in relation to the objects in the recordings, for example a tree, or a fence, etc. The purpose of this is to provide the prosecutors with sufficient supporting evidence in addition to recordings, enhanced images, confessions and other physical kinds of evidence.

Such empirical methods are considered essential for the prosecution to be successful, since in most countries, digital evidence alone has not been readily accepted in the courts of law.

"It is not enough to simply produce an unbiased and technically accurate document describing the outcome of a forensic examination. The primary purpose of the statement is to assist the court in evaluating the admissibility and weight of any evidence found on the digital devices examined for the case" says Ian Kennedy, forensic computer analyst for Kent Police [69].
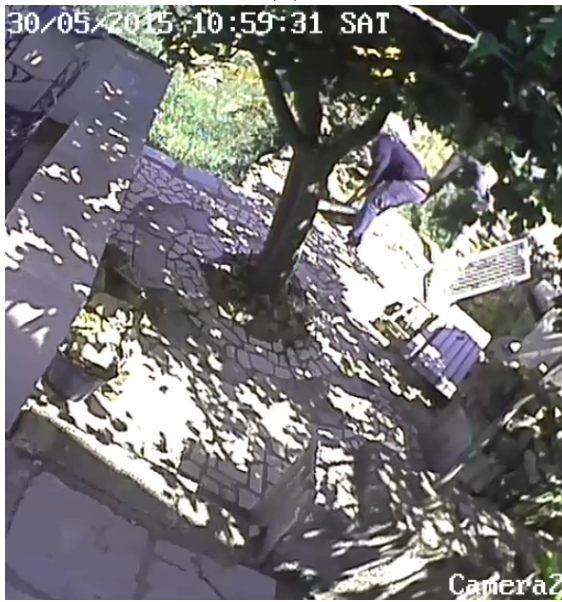
And he continues:"Statements are submitted to the Crown Prosecution Service (CPS) and copies are distributed to both the prosecution and defence counsel. The statement aids

(a)   (b)

(c)   (d)

Figure 10: Day scene. Three young men break into the property and steal the garden gate.

19

Figure 11: Amped FIVE processed image.

the understanding of the examiner's findings and assists in deciding the strategy and the legal points to prove.

Under UK law each offence has what are known as 'points to prove'. For example, under Section 3 of the Computer Misuse Act 1990 a person is guilty of unauthorized modification of computer material if it can be proven that he or she:

- does any act which causes an unauthorized modification of the contents of any computer; and

- at the time when the act was performed he or she has the requisite intent and the requisite knowledge to do so.

These two points demonstrate what in legal terms is called the actus reus (guilty act) and the mens rea (intent/knowledge) of the individual." [69]

Damage to and theft from a property was recorded by Swann View Pro (DVR-4 1260). The video surveillance records were made available to the police and the prosecution, together with the victim's and the witnesses' reports. The officially appointed digital forensic experts carried out video analytics using Amped FIVE [70] video enhancement software, which is "specifically designed for investigative, forensic, military and security operations". Amped FIVE (and similar products currently on the market) provides the chain of filters so that in each step of the process a filter can be applied to deblur an input image or stabilise a video sequence. It thus generates an output that can be passed on to another filter. Some modification of the parameters is possible and experimentation with the chain of filters is also possible in an attempt to align the process of analysis to the features of specific images and video sequences. This software brings the clarification tools for both video and still images into a single package and thus helps the investigators to ascertain that the integrity of the original evidence has been maintained. This is an important consideration since it has been shown that undocumented or modified files have a strong chance of being excluded by the court.

The shortcomings of this video analytic technique are largely due to its lack of adaptability to variations in the outdoor environment, as demonstrated by the use-case scenario discussed in this paper. The environmental conditions at the location where recordings were taken were complex, with large contrast between the sunny and the shaded areas, thus stabilising and deblurring was exceptionally difficult using filter-based approaches as shows the filtered image in Fig. 11. This strongly influenced the prosecutors' decision not to take the matter to court due to insufficient evidence as the police provided only the digital evidence but no supporting empirical evidence was available. The prosecution thus acknowledged that criminal offence was indeed committed but the surveillance videos were insufficient to identify the perpetrators "without any reasonable doubt". The incident was declared to be a "criminal offence by unidentified perpetrators".

5.2.3. Multiple Images Super-resolution as a Forensic Tool

The victim provided us with two videos recorded by camera 2 which last for 410 and 732 seconds, respectively. Those videos were directly exported from the recorder and constitute an accurate and complete replica of the primary videos. From them, the video that showed the assailants' faces, either totally or partially, was extracted. It consists of 10255 colour frames of size 704×576 pixels stored in an MP4 container. Video, in PAL format at 25 frames per second, was compressed using MPEG-4 AVC (H.264) codec at approximately 2.5 Mbps. An audio track was also recorded but no information could be obtained from it. Video compression was not very severe and, although compression always discards detail information from the video sequence, artifacts were not perceptible in most of the scenes with low motion.

Since the goal is to identify the assailants, we extracted from the video those frames where faces were present. Super-resolution, as presented in this paper, is based on the fact that different frames contain complementary information from the scene. According to the model of the registration parameters in Section 3.3, we consider global motion between the different frames, and only displacement and rotations are taken into account. Hence, to obtain good results, super-resolution can not be applied to a series of frames with face in very different poses or with very different sizes. With those constraints in mind, only six frames, from 6762 to 6767 in the video, were initially selected. After frame 6767, a branch covered the face of the assailant, making impossible to extract information from them.

Figure 12a corresponds to frame 6762 and shows one of the assailants with uncovered face. From frame 6762 and five consecutive frames, an area of size 30×19 pixels containing the face was selected, thus obtaining the six images depicted in Fig. 12b that were the initial input of our super-resolution algorithm. Note that the face itself occupies a tiny area of approximately 11×15 pixels.

Several experiments were performed to obtain enhanced versions of the frame 6762 in Fig. 12b. From the six selected input images, those corresponding to frames 6766 and 6767 had to be discarded since the face was partially covered by the branch and it degraded the quality of the output images, so only four input images were considered, see Fig. 13a. For all the used algorithms, given the extremely low number of input frames, the output image size was increased by a factor of 2 in each direction, that is, $P = 4$ was used. Note that four

(a) Frame 6762.

(b) Extracted areas.

Figure 12: (a) Frame 6762 with the face marked. (b) Extracted areas from frames 6762 to 6767.

is a sensible number of frames for $P = 4$. In real-life video, where motion is seldom global and pose changes transform the face images, small warping estimation artifacts may be introduced in the reconstruction process which make the final super-resolved image blurrier. To prevent that from happening, better results are usually obtained by selecting a small number of images showing the same object of interest in very similar poses.

Our first attempt was to use bicubic interpolation. Figure 13b shows the bicubic interpolation of the reference image (frame 6762), depicted in Fig. 12b and the upper left image in Fig. 13a. Interpolation increases the size of the image but it does not increase its resolution, that is, the same information in the low resolution image is present in the interpolated image and the only effect is to smooth pixels boundaries, thus producing an over-smoothed image more pleasant to the visual system but not better defined. We also used the classic method proposed by Molina *et al.* [64] which utilizes a SAR prior model for the image. The resulting image (see Fig. 13c) clearly shows better level of detail than bicubic interpolation but ringing and other artifacts are also present in the image.

The proposed method was also tested using Alg. 1 with the same initial parameters as in the synthetic experiments, except for the integration and blurring matrix $\mathbf{H}_k$ which was set to the convolution matrix obtained from a 3×3 Gaussian kernel of variance 1. The image obtained when using the log penalty function (see Table 1) is depicted in Fig. 13d. It shows a clear improvement in resolution, without the artifacts presented in Fig. 13c. The proposed method was also run using the ℓ_p prior with several values of p. The image obtained for $p = 2$ is displayed in Fig. 13e, which has the same artifact problems as the SAR reconstruction in Fig. 13c. This is expected since the quadratic prior is not able to clearly define the edges of the objects. Using lower values for p allowed us to recover a smooth image without those artifacts (see Figs. 13f–13i). ℓ_p penalty functions with $p \leq 1$ produce very good results when the input images have a very high noise level [12] since they assume piece-wise images, that is, they work well for images with smooth areas separated by strong edges. However, since our video does not contain a significant amount of noise, those priors produce on our images

22

(a) Observations   (b) Bicubic interpolation   (c) Method in [64]

(d) Proposed (log)   (e) Proposed (ℓ_2)   (f) Proposed (ℓ_1, 1 iter.)

(g) Proposed (ℓ_1, 2 iter.)   (h) Proposed ($\ell_{0.8}$, 1 iter.)   (i) Proposed ($\ell_{0.8}$, 2 iter.)
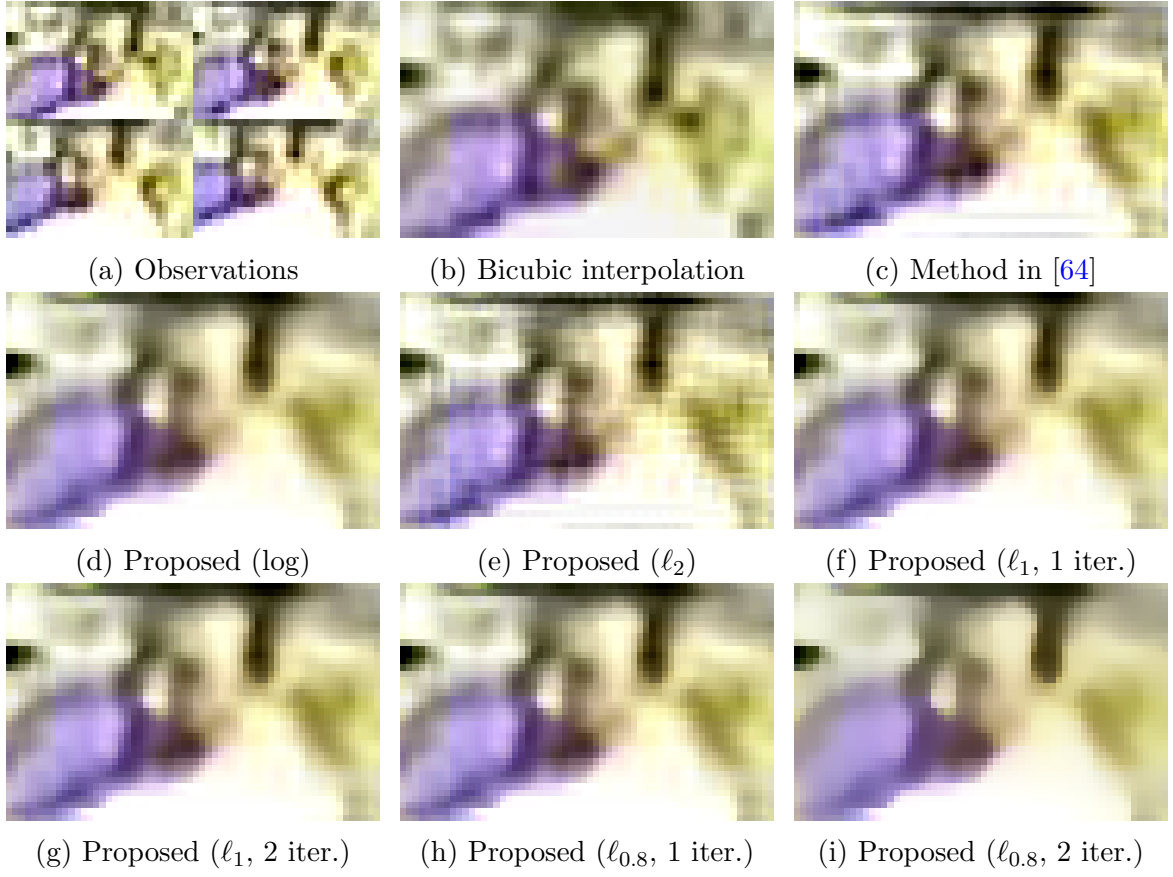
Figure 13: Observations and super-resolved images using different methods.

over-smoothed cartoon-like reconstructions. To ameliorate this effect, the super-resolution method is stopped before convergence. So, Figs. 13f and 13h show the result after one iteration of the SR algorithm in Alg. 1 for $p = 1$ and $p = 0.8$, respectively, while Figs. 13g and 13i depict the corresponding reconstructions with two iterations of the algorithm. They produce over-smoothed areas with sharp edges. The gain in resolution is spoiled by the loss of detail in the figure. The more iterations the algorithm runs, the smoother those areas are.

The original and super-resolved images are included in a detailed expert report. The report has to be complete, accurate, and comprehensive and should be written for the intended audience. The delivered report also includes all the steps required to reproduce the results, ensuring reliability on the process as well as the scientific facts that support the SR method. Following the international guidelines for digital investigations, we could trace how they were actually implemented in the use-case scenario reported here. The correct sequence of steps was followed, but certain aspects of the criminal investigation were neglected, especially regarding the empirical evidence gathering on the part of the police to support the digital evidence presented in the surveillance video.

The proper process was triggered by the victim's report of the crime and the statement describing the incident. Digital evidence was submitted in its original format and the police carried out digital investigation using commercially available software, the two most important steps to ensure the authenticity of the images as evidence. The police sent the outcome of the forensic investigation to the prosecution office as the only evidence, as is often the case when the police do not support forensic analysis with empirical investigations. The prosecution office decided not to pursue the case since the quality of the images was insufficient to unambiguously identify the perpetrators and there was no empirical evidence to compensate for this. Digital evidence alone was deemed sufficient to establish that a criminal offense had been committed but was not sufficient to identify the perpetrators "beyond any reasonable doubt".

It was therefore concluded that the authenticity requirement was met, but the reliability requirement was not, since no other supporting evidence was provided by the police. From the legal perspective, the main issue was the quality of the enhanced images, or more precisely, the enhancement process focused primarily on preserving authenticity of the images produced results that were of insufficient quality. Subsequently, this research team carried out an MISR analysis which produced superior enhanced images, in order to explore whether such images would stand the test of image quality sufficient to establish the reliability of digital evidence. These were the basis of the pilot study, a small-scale investigation to identify the research issues that need to be addressed in the main evaluation of the MISR tools for Digital Forensic Imaging. The study was focused on visual quality evaluation guided by usability methods of the multidisciplinary field of Human-Computer Interaction [71]. Originally these evaluation techniques have been developed to assess and measure user experience of digital media, focused on understanding how visual images influence human perception and cognition[72].

In the pilot study the methods were predominantly qualitative and evaluation interpretative. The sample included ten subjects. In group 1 there were five subjects familiar with

the location and the suspects. They could in principle be regarded as potential witnesses in a court case where digital evidence would be presented to the court. The subjects were asked whether they could recognize any familiar faces in the images enhanced by commercial software (in Fig. 11) and also in the images enhanced by the MISR technique. In group 2 the subjects were not familiar with the content of the images (neither location nor perpetrators) and could in principle be regarded as potential jury members. They were asked to compare the enhanced images with photographs of non-suspects taken by ordinary cameras, including smart phones. In addition, both groups were asked to compare photographs of several suspects and identify the ones shown in the enhanced images.

The outcomes of the pilot evaluation show clearly that the subjects in Group 1 were able to identify the perpetrators more accurately from the MISR images. Group 2 expressed greater certainty when comparing the MISR images with photographs, than in similar comparisons between commercially enhanced images and photographs. The results were thus overwhelmingly in favour of the MISR enhanced images.

It is difficult to generalize these outcomes without further research involving a larger data set of enhanced images, a larger and more representative sample of informants to include also the legal and forensic professionals involved in the legal process. The application of MISR is therefore research in progress where the multi-disciplinary approach described in this paper is also providing an opportunity for the development of novel research methods to capture both image-oriented and people-oriented perspectives in a uniform analytical framework.

Last but not least, there is significant scientific potential of this work. It is possible to develop an experimental method to study the effect of MISR imaging on human perception and cognition in general. It is also possible to study this effect in particular, for example in the context of specific tasks, such as facial identification, thus addressing the concerns in the digital forensics community to improve reliability of digital evidence.

6. Conclusions

In this paper we have explored the application of imaging science in digital forensics. More concretely, we explored the use of super-resolution techniques, and proved their applicability as a powerful tool for Digital Forensic Imaging able to extract images from a surveillance video sequence with enhanced level of detail whilst preserving the authenticity, credibility, and reliability of video data as digital evidence.

The proposed variational Bayesian approach with super-Gaussian priors provides a much clearer picture than interpolation and classical super-resolution approaches both on synthetic video sequences and the real-life case under study. This result was corroborated by visual inspection of the processed images and, in the case of synthetically generated sequences, also using objective measures.

Illustrated with a real-life case of damages and stealing in a private property, we have addressed all the steps from police report to bringing the case to the court of law. We showed that MISR can enhance frames extracted from an extremely low quality video. However, although the groups of informants identified the suspect in MISR pictures, the quality of the images was insufficient to unambiguously identify the perpetrators. Obviously, analysis

of video sequences taken in adverse outdoor conditions, such as those in this case study, cannot be regarded as sole evidence in a legal process. In fact, our case study represents a worst-case scenario from the perspective of the analysis of digital evidence, but in real life, images of such poor quality are not uncommon.

Traditional police methods are not replaced by the analysis of digital evidence, rather, they are enhanced by techniques such as MISR presented in this paper. Image analysis is only one of the available methods for processing digital evidence, in successful prosecutions. They were combined with empirical methods, such as metrics to establish the perpetrator's height in relation to objects on the crime scene such as trees, fences, etc.

There are therefore considerable advantages of the MISR approach to image enhancement in real-life digital forensics contexts, and these are explored in our research in progress. In addition, this kind of multi-disciplinary approach can contribute to the broader scientific study of the effects of digital images on human perception and cognition than is currently the case.

References

[1] B. D. Carrier, Defining digital forensic examination and analysis tools, https://www.dfrws.org/conferences/dfrws-usa-2002/sessions/defining-digital-forensic-examination-and-analysis-tools (2001).

[2] B. D. Carrier, Basic digital forensic investigation concepts, https://web.archive.org/web/20100226184652/http:/www.digital-evidence.org/di_basics.html (Jun. 2006).

[3] E. Casey, Handbook of digital forensics and investigation, Academic Press, 2009.

[4] E. Casey, Digital evidence and computer crime: Forensic science, computers, and the internet, Academic Press, 2011.

[5] M. Reith, C. Carr, G. Gunsch, An examination of digital forensic models, International Journal of Digital Evidence 1 (3) (2002) 1–12.

[6] S. L. Garfinkel, Digital forensics research: The next 10 years, Digital Investigation 7 (2010) S64–S73.

[7] J. Williams, ACPO Good Practice Guide for Digital Evidence, http://www.digital-detective.net/digital-forensics-documents/ACPO_Good_Practice_Guide_for_Digital_Evidence_v5.pdf.

[8] National Institute of Justice, Forensic examination of digital evidence: A guide for law enforcement, https://www.ncjrs.gov/pdffiles1/nij/199408.pdf (Apr. 2004).

[9] N. A. Alfaize, The impact of culture and religion on digital forensics: The study of the role of digital evidence in the legal process in Saudi Arabia, Ph.D. thesis, Faculty of Technology, De Montfort University, Leicester, UK (2015).

[10] S. D. Babacan, R. Molina, A. Katsaggelos, Variational Bayesian super resolution, IEEE Transactions on Image Processing 20 (4) (2011) 984–999.

[11] S. Villena, M. Vega, D. Babacan, R. Molina, A. Katsaggelos, Bayesian combination of sparse and non sparse priors in image super resolution, Digital Signal Processing 23 (2) (2013) 530–541.

[12] S. Villena, M. Vega, R. Molina, A. Katsaggelos, A non-stationary image prior combination in super-resolution, Digital Signal Processing 32 (2014) 1–10.

[13] J. A. Palmer, K. Kreutz-Delgado, S. Makeig, Strong Sub- and Super-Gaussianity, in: Latent Variable Analysis and Signal Separation: 9th International Conference, LVA/ICA, Springer Berlin Heidelberg, 2010, pp. 303–310.

[14] S. Babacan, R. Molina, M. Do, A. Katsaggelos, Bayesian blind deconvolution with general sparse image priors, in: European Conference on Computer Vision (ECCV), 2012, pp. 341–355.

[15] M. Vega, R. Molina, A. K. Katsaggelos, Parameter estimation in Bayesian blind deconvolution with super Gaussian image priors, in: Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European, 2014, pp. 1632–1636.

[16] X. Zhou, M. Vega, F. Zhou, R. Molina, A. K. Katsaggelos, Fast Bayesian blind deconvolution with Huber Super Gaussian priors, Digital Signal Processing 60 (2017) 122–133.

[17] R. Saferstein, Criminalistics: An Introduction to Forensic Science, 11th Edition, Pearson, 2009.

[18] ISO/IEC 27037:2012 – Information technology – Security techniques – Guidelines for identification, collection, acquisition, and preservation of digital evidence, https://www.iso.org/standard/44381.html (2012).

[19] ISO/IEC 27042:2015 – Information technology – Security techniques – Guidelines for the analysis and interpretation of digital evidence, https://www.iso.org/standard/44406.html (2015).

[20] E. Berg, Legal ramifications of digital imaging in law enforcement, Forensic Science Communications 2 (4) (2000) 1–19.

[21] S. H. James, J. J. Nordby, S. Bell, Forensic Science: An Introduction to Scientific and Investigative Techniques, 4th Edition, CRC Press, 2014.

[22] A. E. Dirik, A. Karaküçük, Forensic use of photo response non-uniformity of imaging sensors and a counter method, Opt. Express 22 (1) (2014) 470–482.

[23] S. D. Anderson, Digital image analysis: Analytical framework for authenticating digital images, Ph.D. thesis, University of Colorado at Denver (2011).

[24] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, S. Tubaro, An overview on video forensics, APSIPA Transactions on Signal and Information Processing 1. doi:10.1017/ATSIP.2012.2.

[25] A. Gehani, J. Reif, Super-resolution video analysis for forensic investigations, Advances in Digital Forensics III (2007) 281–299.

[26] A. K. Katsaggelos, R. Molina, J. Mateos, Super Resolution of Images and Video, Morgan and Claypool, 2007.

[27] P. Milanfar, Super-Resolution Imaging, Digital Imaging and Computer Vision, Taylor & Francis / CRC Press, 2010.

[28] K. Nelson, A. Bhatti, S. Nahavandi, Performance evaluation of multi-frame super-resolution algorithms, in: International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2012, pp. 1–8.

[29] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, L. Zhang, Image super-resolution: The techniques, applications, and future, Signal Processing 128 (Supplement C) (2016) 389–408.

[30] C.-Y. Yang, C. Ma, M.-H. Yang, Single-image super-resolution: A benchmark, in: European Conference on Computer Vision, Springer, 2014, pp. 372–386.

[31] L. Zhang, X. Wu, An edge-guided image interpolation algorithm via directional filtering and data fusion, IEEE Transactions on Image Processing 15 (8) (2006) 2226–2238.

[32] L.-J. Deng, W. Guo, T.-Z. Huang, Single image super-resolution by approximated Heaviside functions, Information Sciences 348 (2016) 107–123.

[33] P. Rasti, O. Orlova, G. Tamberg, C. Ozcinar, K. Nasrollahi, T. B. Moeslund, G. Anbarjafari, Improved interpolation kernels for super resolution algorithms, in: Image Processing Theory Tools and Applications (IPTA), 2016 6th International Conference on, IEEE, 2016, pp. 1–6.

[34] W. Yang, J. Feng, J. Yang, F. Zhao, J. Liu, Z. Guo, S. Yan, Deep edge guided recurrent residual learning for image super-resolution, IEEE Transactions on Image Processing 26 (12) (2017) 5895–5907.

[35] X. Lu, Y. Yuan, P. Yan, Image super-resolution via double sparsity regularized manifold learning, IEEE Transactions on Circuits and Systems for Video Technology 23 (12) (2013) 2022–2033.

[36] S. Yu, W. Kang, S. Ko, J. Paik, Single image super-resolution using locally adaptive multiple linear regression, Journal of the Optical Society of America A 32 (12) (2015) 2264–2275.

[37] C. Dong, C. C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (2) (2016) 295–307.

[38] P. Rasti, K. Nasrollahi, O. Orlova, G. Tamberg, T. B. Moeslund, G. Anbarjafari, Reducible dictionaries for single image super-resolution based on patch matching and mean shifting, Journal of Electronic Imaging 26 (2) (2017) 023024–023024.

[39] Y. Romano, J. Isidoro, P. Milanfar, RAISR: rapid and accurate image super resolution, IEEE Transactions on Computational Imaging 3 (1) (2017) 110–125.

[40] A. Lucas, M. Iliadis, R. Molina, A. Katsaggelos, Using deep neural networks for inverse problems in imaging, Signal Processing Magazine, (accepted for publication).

[41] S. Villena, Superresolucion y reconstrucción Bayesiana de imágenes a partir de imágenes de baja resolución rotadas y desplazadas. Combinación de modelos, Ph.D. thesis, Universidad de Granada (Dec. 2011).

[42] Y. Zhu, K. Li, J. Jiang, Video super-resolution based on automatic key-frame selection and feature-guided variational optical flow, Signal Processing: Image Communication 29 (8) (2014) 875–886.

[43] K. Li, Y. Zhu, J. Yang, J. Jiang, Video super-resolution using an adaptive superpixel-guided auto-regressive model, Pattern Recognition 51 (2016) 59–71.

[44] E. Faramarzi, D. Rajan, F. C. Fernandes, M. P. Christensen, Blind super resolution of real-life video sequences, IEEE Transactions on Image Processing 25 (4) (2016) 1544–1555.

[45] V. K. Banse, K. Chandravanshi, A novel approach for super resolution of video frame using spatially adaptive total variation, International Journal of Computer Applications 176 (4) (2017) 29–34.

[46] R. Molina, A. Katsaggelos, L. Alvarez, J. Mateos, Towards a new video compression scheme using super-resolution, in: J. G. Apostolopoulos, A. Said (Eds.), Proc. of SPIE-IS&T Electronic Imaging, Visual Communications and Image Processing 2006, Vol. 6077, 2006, pp. 607706–1 607706–13.

[47] L. Zhang, H. Zhang, H. Shen, P. Li, A super-resolution reconstruction algorithm for surveillance images, Signal Processing 90 (3) (2010) 848–859.

[48] A. Piva, An overview on image forensics, ISRN Signal Processing 2013.

[49] J. Satiro, K. Nasrollahi, P. Correia, T. B. Moeslund, Super-resolution of facial images in forensics scenarios, in: IEEE 5th International Conference on Image Processing Theory, Tools and Applications, 2015, pp. 55–60.

[50] T. Uiboupin, P. Rasti, G. Anbarjafari, H. Demirel, Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring, in: Signal Processing and Communication Application Conference (SIU), 2016 24th, IEEE, 2016, pp. 437–440.

[51] J. Shao, F. Chao, M. Luo, J. C. Lin, A super-resolution reconstruction algorithm for surveillance video, Journal of Forensic Science and Medicine 3 (2017) 26–30.

[52] S. A. Ledesma, A proposed framework for forensic image enhancement, University of Colorado at Denver, 2015.

[53] E. Verolme, A. Mieremet, Application of forensic image analysis in accident investigations, Forensic Science International 278 (2017) 137–147.

[54] E. S. Lee, M. G. Kang, Regularized adaptive high-resolution image reconstruction considering inaccurate subpixel registration, IEEE Transactions on Image Processing 12 (7) (2003) 826–837.

[55] M. E. Tipping, C. M. Bishop, Bayesian image super-resolution, in: Advances in Neural Information Processing Systems 15 (NIPS), MIT Press, 2003, pp. 1303–1310.

[56] L. C. Pickup, D. P. Capel, S. J. Roberts, A. Zisserman, Bayesian methods for image super-resolution, The Computer Journal 52 (1) (2009) 101–113.

[57] A. Kanemura, S.-I. Maeda, S. Ishii, Superresolution with compound Markov random fields via the variational EM algorithm, Neural Networks 22 (7) (2009) 1025–1034.

[58] T. Katsuki, A. Torii, M. Inoue, Posterior-mean super-resolution with a causal Gaussian Markov random field prior, IEEE Transactions on Image Processing 21 (7) (2012) 3182–3193.

[59] R. T. Rockafellar, Convex analysis, Princeton University Press, 2015.

[60] B. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: Proceedings of Imaging Understanding Workshop, 1981, pp. 121–130.

[61] C. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.

[62] Z. Chen, S. D. Babacan, R. Molina, A. K. Katsaggelos, Variational bayesian methods for multimedia problems, IEEE Transaction on Multimedia 16 (4) (2014) 1000–10017.

[63] M. Beal, Variational algorithms for approximate Bayesian inference, Ph.D. thesis, The Gatsby Computational Neuroscience Unit, University College London (2003).

[64] R. Molina, M. Vega, J. Abad, A. Katsaggelos, Parameter estimation in Bayesian high-resolution image reconstruction with multisensors, IEEE Transactions on Image Processing 12 (12) (2003) 1655–1667.

[65] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.

[66] D. Chawla, M. C. Trivedi, A comparative study on face detection techniques for security surveillance, in: S. Bhatia, K. Mishra, S. Tiwari, V. Singh (Eds.), Advances in Computer and Computational Sciences, Vol. 554 of Advances in Intelligent Systems and Computing, Springer, Singapore, 2018, pp. 531–541.

[67] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Vol. 1, 2001, pp. I–511–I–518.

[68] L. Alvarez, J. Mateos, R. Molina, A. Katsaggelos, High resolution images from compressed low resolution video: Motion estimation and observable pixels, International Journal of Imaging Systems and Technology 14 (2) (2004) 58–66.

[69] I. Kennedy, Presenting digital evidence to court, http://www.bcs.org/content/ConWebDoc/7372 (Oct. 2006).

[70] Amped Software, Amped FIVE: Forensic image and video enhancement, https://ampedsoftware.com/five (2017).

[71] G. Cockton, Usability evaluation, in: M. Soegaard, R. F. Dam (Eds.), The Encyclopedia of Human-Computer Interaction, 2nd Edition, The Interaction Design Foundation, 2013, Ch. 15.
URL https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction/usability-evaluation

[72] E. Law, V. Roto, M. Hassenzahl, A. Vermeeren, J. Kort, Understanding, scoping and defining user experience: A survey approach, in: Proceedings of Human Factors in Computing Systems conference, CHI'09, 2009, pp. 719–728.