ESTIMATION OF FRAME SEQUENCE NOISE WITH REMOVAL OF JPEG ARTIFACTS

VASILEIOS I. ANAGNOSTOPOULOS∗, EMMANUEL S. SARDIS† and THEODORA A. VARVARIGOU‡

Institute of Communications and Computer Science
National Technical University of Athens-NTUA
9, Iroon Polyteknion Street
Zografou 15773, Athens, Hellas, Greece
∗vanag@mail.ntua.gr
†sardism@mail.ntua.gr
‡dora@telecom.ntua.gr

Received 12 July 2010
Revised 11 December 2012
Accepted 12 December 2012
Published 26 April 2013

This paper proposes a method to remove JPEG noise artifacts from frame sequences. Using extensive experimental results we show how an online system with periodic noise estimation functionality can estimate the real frame noise even if the images are in JPEG format. We present the mathematical basis of the methodology and show in real content that we can have reliable measurements. We also present the results obtained on a real network camera and show that our method can provide a much better estimation of the noise standard deviation compared to common practice but comparable inter-channel and spatial intra-channel correlation estimates. We also provide some guidelines for capturing datasets necessary to apply computer vision tasks. Our approach exploits the well known stochastic linearization phenomenon which we prove that is present in our case.

Keywords: Noise; image processing; image restoration.

1. Introduction

CCD (Charged Coupled Device) camera noise estimation and de-noising or noise reduction in video sequences is a very old subject following the development of computer vision (see Ref. 1). It lies at the heart of signal processing and computer vision since it gives limits on the performance of a sensor by providing fundamental bounds on signal recovery. The well-known SNR figure of merit finds great applicability in areas ranging from photonics to edge detection in computer vision and has been used for years as an indicator of receptor quality. Central to this
approach is the modeling of noise in the form of a Gaussian process which in many cases permits closed form solutions for typical problems. In practical applications, the variance is not known \textit{a priori} and it is estimated from the data given some reasonable assumptions (see Ref. 5, Chap. 3; and Ref. 6). Given this information, parameter tuning on the system like an increase in transmitter power can be accomplished in order to match desired performance characteristics. The Gaussian noise model is not the only one and in the context of cameras and other electronic imaging systems other noise models are more appropriate. A very interesting point in these approaches is that the variance is no longer independent from image channel intensity contrary to the common sense. This subtle point gives rise to interesting phenomena.

Computer vision with its numerous applications has gained significant momentum the last decade mainly due to the increase in processing power with the introduction of improved computer systems. The same improvements have made possible the variety of imaging systems for consumer usage usually embedded in consumer electronics. The number of digital photographs available in the Internet increases rapidly which is the sign of the importance of visual content in our everyday life. Still problems exist owing to both the nature of the camera systems and the compression methods used to reduce the bandwidth requirements. It is still impossible to match the image quality offered by the human vision system though particular applications for these devices have flourished. Various tracking or detection algorithms that process this overwhelming content are very common in literature these days and find practical applications. They fail in many cases owing to the poor image quality originating from device impairments or the very common JPEG compression scheme both of which invalidate the basic assumptions behind these efforts. Reference 8 agrees on this point. Due to simplifications for the sake of tractability the existing body of work rarely makes any real-life assumptions on the capture device. Instead researchers are focusing on image understanding rather than modeling the image acquisition and compression process which is an important subject. Many of them assume the generic white noise models or try to reduce illumination variation effects even if they are attributed to the camera response functions. A realistic model of imaging devices and JPEG compression artefacts can help improve the characteristics of existing algorithms and even motivate the development of new approaches with better performance. This is not unreasonable because researchers can gain better insight in the functionality of their algorithms. The complexity of computer vision algorithms is usually very high and proper understanding is limited to qualitative explanations. Modeling noise is a first step towards quantitative conclusions. These conclusions even in numerical form, due to the intractability of the corresponding mathematics, are vital to the boost of the exciting research taking place in this challenging scientific area. Our contribution helps in this area by providing a method to remove JPEG compression artifacts from camera noise estimation by using a compensation scheme suitable for real time applications.
2. CCD Noise Modeling for Image Sensors

In this section, we will briefly touch upon the main sources of noise in digital cameras. We will not delve into details because there are excellent references which describe the process of image generation thoroughly.\textsuperscript{7,9,10} We will isolate key processes for our purpose. This section is mainly a review with the intent to fix notation and collect fundamental existing knowledge upon which we will build our work. Our description is simple though it accounts for a high-level description necessary for the computer vision scientist to understand the sensor device prior to video acquisition. In practice the pipeline is unknown due to the proprietary nature of the imaging systems which are trade secrets and under-documented. As we will see, it is a very serious, challenging task of particular interest for the image processing community to reverse engineer the parameters of the model. We restrict our attention to CCD cameras because they are the predominant imaging devices and because our experimental results are captured with such a camera. Each pixel can be modeled with a capacitor connected to a current source, the photo-detector. The capacitor is charged for some specific time $T$ called the exposure time. Care is taken to select an exposure time sufficient to keep the capacitor far from saturation with the intent of having an approximately linear relation between the charge and the photo-current. At the end of the exposure time the sensor samples the voltage of the capacitor and the capacitor is subsequently discharged in order to start a new sampling period. There are already two sources of noise here. The first one is attributed to the shot noise effect mainly due to the discrete nature of light which is a pure quantum effect. For this reason the accumulated charge can be modeled with a Poisson process with mean value proportional to the incident electric field. The second noise is the well-known thermal noise in the electronic circuitry. It can be modeled with a Gaussian process. Each voltage sample is presented to an amplifier. In CMOS devices the gain of the amplifier can have significant inter-pixel variation owing to the difference in operating point and manufacturing processes. In CCD devices there is typically only one amplifier (in some cases there is column-wise or row-wise bank of amplifiers,\textsuperscript{11} for CMOS\textsuperscript{12}) and the CCD values are scanned using a combination of shift registers before amplified.

However, there is still variance on the operating point due to thermal fluctuations. This phenomenon gives rise to the statistical component of gain pattern noise.\textsuperscript{13} The fixed component of gain pattern noise can be attributed to the difference on the electronic elements not including the amplifier in case of single amplifier CCD. We should point out here that the amplifier is not linear in general and consequently the modeling of the amplifier with a constant gain is approximate. In literature, some researchers\textsuperscript{10,14} indicate that before quantization a nonlinear function is applied to the amplifier output with the intent of correcting the nonlinearity of previous stages. Typically this nonlinearity takes the form of a power law with exponent of value 2.2 (NTSC) or 2.8 (PAL). However, there is no sufficient evidence as to why this is necessary in MJPEG (Motion-JPEG) streams. It can be regarded as a
rough camera response function estimate but there is excellent work on this topic\textsuperscript{15} and we avoid making such an assumption. For presentation to various devices like screens or printers this nonlinear correction, usually called “gamma-correction”, is necessary and it is accomplished by the windowing architecture of the operating system. Therefore it is not necessary to pre-store this value. For a CCD camera such correction would make hardware more complex and would be inflexible. Moreover in our tests no such power law is necessary to interpret experimental data. Finally the voltage values are quantized for storage reduction. In 3CCD cameras which are usually very expensive this is a typical pipe-line used for the formation of frames in digital cameras. In cheaper cameras there is a de-mosaic step before the quantizer due to the Bayer pattern used in CC\textsuperscript{D}s.\textsuperscript{16}

Figure 1 shows the mathematical description of the CCD pipeline. The quantity of the system corresponding to a value in the pixel frame buffer of the camera is $I_4$. Noise source $N$ can be modeled as a Gaussian with nonzero mean owing to the dark current $I_0$ and variance $\sigma$.

The $I_1$ signal is a scalar multiple (due to the capacitor) of a Poisson distribution. Since the number of generated electrons is very big, signal $I_1$ can be modelled with a Gaussian distribution\textsuperscript{4} with mean $I$ and variance proportional to $I$. The mathematical quantity $I$ is proportional to the incident energy shown as $E$. For this reason, signal $I_2$ can be modelled with a Gaussian distribution with mean $I + I_0$ and variance equal to an affine function of $I$. The amplifier can be modelled as a linear device with input output relation

$$I_3 = AI_2 + B, \quad (1)$$

where in Eq. (1), $A$ is its gain and $B$ is the amplifier offset which is usually neglected. The last assumption is not unusual since due to the linear nature of the amplifier, the offset can be absorbed by the dark-current. Finally the quantization is applied to derive the pixel-values in the frame-buffer. To this end the above chain of operations

![Fig. 1. CCD pipeline.](image)
results to the following model, Eqs. (2) and (3) for the stochastic nature of a pixel at \( x-y \) coordinates.

\[
I_4(x, y) = Q(I_3(x, y)), \tag{2}
\]

\[
I_3(x, y) = A(x, y)I_1(x, y) + A(x, y)N(x, y) + B(x, y). \tag{3}
\]

For Eq. (2), \( Q(\cdot) \) is the quantizer function that quantizes its argument to an integer in the range 0–255. We usually refactor the gain as a constant multiplied with a Gaussian random variable with mean value 1 and variance \( \sigma_A \) owing to the temporal variance of the gain.\(^{17} \) We also assume that the offset is a Gaussian random variable with mean 0 and variance \( \sigma_B \). Typical applications remove from the mean value the term due to the dark current with a simple differencing scheme. This scheme is implemented in practice with shooting frames in a dark place and subtracting the mean values of the pixels from every frame before processing. This naive model carries certain disadvantages:

(i) The noise before the amplifier is approximately Gaussian with variance dependent on signal strength. In reality it is Poisson.

(ii) The amplifier is not necessarily linear.

(iii) Inter-channel correlation arises partly due to the nature of the images, and partly due to cross-talk.

(iv) The existence of the quantizer changes the noise characteristics especially at low and high values.

However, this model is used repeatedly in literature because of its simplicity. Even in this simple form it creates difficult problems in typical tasks like the Canny edge detector\(^{18} \) that assumes constant variance. To see this, the variance of the above formula takes the well known form (in a region with constant color \( I \)).

\[
\text{Var}\{I_3(x, y)\} = A^2\{\sigma_A^2(I^2 + \lambda I + \sigma^2 + I_0^2) + \lambda I + \sigma^2\} + \sigma_B^2, \tag{4}
\]

which is called the CCD sensor Eq. (10). In the big study on feature descriptors\(^{19} \) this form of noise is not taken into account which is a serious omission if real-world performance is to be expected. We point out again that the above equation deviates from the typical assumed form of constant variance. A possible explanation for this omission is that the above mentioned references try to be generic since the above noise model is camera dependent and typical values are not readily available.

### 2.1. The effect of compression and prior work

The above chain of operations is usually sufficient to describe newer cameras like those supporting the GiGE Vision standard.\(^{20} \) Typically there is another source of distortion which is of paramount importance to every computer vision system and usually neglected. An online system cannot grab frames in real-time without some form of compression of the data. There are a growing number of products giving
raw (uncompressed) frames for reliable computer vision applications. The emerging GiGE Vision cameras allow for applications that can process frames at 25 FPS or more without any compression but require 1000 MBps Ethernet networks for operation which can be impractical due to the big amount of data and the container format of the IP protocol. The manufacturers of these cameras assume that these cameras can be operated in isolated Gigabit Ethernet LANs in an IEEE1394 way. On the other hand there is an existing range of solutions which provide MJPEG streams over HTTP via an open API, something that makes them particular attractive for truly cross-platform and vendor independent integration to computer vision systems. The big benefit is the multiplexing of many streams in a 1000 MBps network instead of only one stream. This a very attractive option because it makes multi-camera setups more affordable and versatile. Reuse of existing Ethernet networks for transport allows easy deployment in different environments and a simplified architecture. Moreover the adaptation of the quality of the stream to the networking conditions can be crucial. For this reason, JPEG distortion, usually treated as content dependent noise, is necessary for improved performance of computer vision algorithms. However, the effects of JPEG compression are evaluated visually and not in real-time systems with few exceptions using a software post-processing solution.

JPEG artifact removal for frame sequences has not received the focus it deserves from the computer vision society. As the technology matures we believe it will become minor problem since frames will be transmitted uncompressed. However, it is not something that will happen soon. Until then work is necessary to ensure reliability of current systems. In our work we intend to facilitate the creation of easy to use distortion models applicable in real time situations and remove as much as possible compression artifacts. In general current practice has given an extensive set of methods for de-noising images under different assumptions. But it has not focused on the real nature of the noise. As such, these models are content-dependent and do not reflect the true nature of the noise which is camera dependent. An excellent research touching exclusively the problem of camera noise modeling mostly on static images. However, it fails to present a complete and clear picture of compression artifacts. In the same spirit, in Ref. 26, the authors design an edge detector with the assumption of a Skellam distribution on pixel value difference. However, not only does it not account for the aforementioned problems, it still does not account for gain fixed pattern noise and its coupling with compression. Reference 27 accounts for gain fixed pattern noise but still does not take into account compression artifacts. Reference 28 is closer to the spirit of our work because it uses a video sequence to implement a simultaneous noise estimation and de-noising scheme however, it does not explain how to deal with JPEG compression artifacts or why the noise can be Gaussian. The work there can be complemented by ours to offer better results. We also offer complementary data to Ref. 10 in case of video sequences. The big problem of video sequences is that it is difficult to estimate noise because objects move. One has to
do noise estimation and object segmentation simultaneously. The problem becomes very complicated if the noise is not modeled or not accounted for. One must first apply a suitable noise model and then execute the object segmentation step while applying if necessary an estimation of the model parameters. This increases the reliability of tracking or optical flow estimation significantly. One of the problems of these approaches is that they are offline and in reality the noise is content dependent. We offer a compensation scheme that not only makes noise content independent when the images presented to the lossy compressor are corrupted by such noise but also we make correct estimations as far as noise intensity is concerned. In this respect we offer an improved method to model camera specific noise on the acquired images. We also offer an opportunity for a real time system to do correct re-estimation of the noise since we do not require uncompressed images. Noise compensation restores the white nature of the noise (when applicable) and allows the use of the existing signal processing artillery in coping with the problem of reliable pixel change identification. On the other hand, independence from content allows for infrequent model updates, usually owing to aging, sensor heating and parameter tuning. We also remark that our approach can help work done on single image de-noising and restoration in Refs. 25 and 26.

2.2. JPEG distortion modeling

The JPEG standard provides a well-documented means for lossy image compression that parametrizes information loss via a quality number and has acceptable visual distortion of the images. It is the most common means of image compression of Internet static image content and comes with optimized libraries that can compress and decompress content in real-time. Readily available implementations make it attractive for hardware implementations and for this reason many camera capture devices use it almost exclusively along with MPEG2 compression. Though MPEG2 compression results to reduced file sizes and the ability to stream content via the Web with minimal overhead and minimal traffic, network camera manufacturers have always an MJPEG implementation for applications requiring better quality and because of the simplicity of HTTP transport in LAN (local area network) environments. The network overhead is bigger but for computer vision applications it is advantageous to use high quality content. The distortion impacts static images in a tolerable way for humans. For computer vision systems the distortion is not very acceptable\(^{19}\) and for this reason a JPEG quality factor more than 80% is desired. In the field of camera noise modeling most studies arbitrarily ignore this phenomenon. We believe that this is a very bad practice because every stage of the pipeline has noticeable noise side effects. We detail the impacts of every stage:

(i) The YCbCr transformation induces inter-channel correlation.
(ii) Chroma sub-sampling distorts signal.
(iii) DCT (Discrete Cosine Transform) quantization injects noise which after the inversion creates spatial intra-channel correlation.
(iv) The inverse YCbCr induces further inter-channel and spatial correlation (due to sub-sampling stage).

We can identify two processes in the JPEG pipeline. A linear process related to color-space conversions and sub-sampling and the nonlinear process of coefficient quantization. The most problematic step is the highly nonlinear step of quantization of DCT coefficients. A naive approach would be to model the quantization effect by the well-known model of uniform noise. This is not possible. There are no guarantees that the histograms of the coefficients are uniform. For this reason, the classical model of the quantizer as a uniform noise source with variance equal to $\frac{1}{12}$ does not hold. The problem in such a case can be tackled by computing the histogram of coefficients for specific position in the block and model the error. However, this approach distorts finer details of the image.$^{10,29}$ Fortunately in the case of noise measurements the problem is much different. Here the Gaussian noise of the input data is altered by the JPEG distortion. Great care must be taken to clear the measurements from this effect in order to estimate the noise parameters prior to quantization. In the next section, we will try to understand how different is the problem and what can be done. We target exclusively the nonlinear effect.

2.3. Modeling quantizer distortion

A common misunderstanding is the fact that a uniform quantizer distorts a signal with uniform noise. This conclusion holds true under the assumption that the values behind the quantizer are distributed uniformly along its range. Even in this simple case when the quantizer is lower and upper bounded the typical model breaks down at the ends of the input range. In the case of DCT coefficients, quantization is not a serious problem because the coefficients are encoded as 16-bit integers. We readily see that the effect of corner cases is greatly reduced. When a Gaussian variable is quantized we observe that for a broad range of values, the output of the quantizer has mean and variance very close to the corresponding properties of the Gaussian variable on which it operates. Mathematically this assertion is very difficult to derive and for this reason we resort to detailed diagrams. Here a uniform quantizer is applied on a Gaussian random variable. The quantizer has 256 levels, which is typical for 8-bit imaging applications, where integers are represented in the range 0-255. Two graphs are presented in Fig. 2. The first one shows the absolute difference between output and input mean value of the quantizer for fixed variance values greater than 2 and the second ratio of the output standard deviation to the input standard deviation of the quantizer for fixed variance values greater than 2 again. It is easy to observe the great stability of the quantizer. The mathematical formulae behind these figures are presented in the Appendix.

In our measurements, ranges of variance between 0.1 and 0.4 were also recorded. This is the most difficult region to model. Although the mean values are very stable, the measured variance has a strong dependence on the input mean. Before we offer
Estimation of Frame Sequence Noise with Removal of JPEG Artifacts

Fig. 2. Absolute difference between output and input mean (left), ratio of output standard deviation to input standard deviation (right).

Fig. 3. Oscillations in the output mean (left) and output standard deviation (right) in the low noise region.

an explanation of the phenomenon we display the observed fluctuations in this interesting region in Fig. 3 for values 0.1, 0.2, and 0.4 for standard deviation.

We observe that there is a periodic fluctuation of standard deviation with maximum at half integral values and minimum at integral values. In the next figure, (Fig. 4) we record both extreme values of the oscillation of the standard deviation for the region 0 to 1.6.

It is time to explain the appearance of oscillations. The fact that minimum values are obtained on integral values, while the maximum values are obtained at
half-integral values is not difficult to understand. At integral values of the input mean the error is small due to small perturbations around it which are absorbed by the rounding process. However, at half integral values, great instability is observed because small deviations push the quantizer to neighboring integral values and for this reason they are enhanced. This can be explained easily quantitatively. For input standard deviation smaller than 0.3 and mean value at 128.5 the fluctuations push the value at three times the sigma value right or left of the mean. This means that due to quantization values of 128 and 129 appear most of the times equi-probably in the output of the quantizer. Consequently we can compute a standard deviation approximately equal to 0.5 which can be seen from the above diagrams. It is not difficult also to see that approximately the output of the quantizer can be seen as a Gaussian process with mean value $\mu_{\text{output}}$ and standard deviation $\sigma_{\text{output}}$ related by a functional relation of the form,

$$ (\mu_{\text{output}}, \sigma_{\text{output}}) = (f_1(\mu_{\text{input}}, \sigma_{\text{input}}), f_2(\mu_{\text{input}}, \sigma_{\text{input}})). $$

(5)

The existence of the stability of the mean means that $\mu_{\text{output}} \simeq \mu_{\text{input}}$. Consequently, the only functional relationship of concern is

$$ \sigma_{\text{output}} = f_2(\mu_{\text{output}}, \sigma_{\text{input}}). $$

(6)

We choose for safety a value of 2 instead of 0.6 for standard deviation as the critical value where oscillations vanish (please see Fig. 4). Figure 3 then shows that due to symmetry there is dependence on

$$ g(\mu_{\text{output}}) = \min\{\mu_{\text{output}}, 1 - \mu_{\text{output}}\}. $$

(7)
Estimation of Frame Sequence Noise with Removal of JPEG Artifacts

rather than \( \mu_{\text{output}} \). In Eq. (7), \( \{x\} \) is the fractional part of \( x \). For this reason, there is a function \( f \), such that

\[
\sigma_{\text{output}} \approx \begin{cases} 
 f(g(\mu_{\text{output}}, \sigma_{\text{input}})) & \sigma_{\text{input}} \leq 2 \\
 \sigma_{\text{input}} & \sigma_{\text{input}} \geq 2 
\end{cases}
\] (8)

Given function \( f(\cdot) \), the approximate stability of the mean and the set of Eqs. (7) and (8) allow us to compute the input standard deviation from the output standard deviation with the help of a lookup table. References 30 and 31 tackle the problem in a more general setting. However, we exploit the fact that image sequences are given and the nature of the oscillations to derive simplified equations. The results of this general theory are not used for JPEG artifacts’ removal which is novel in our case but the general theory presented there is applicable even in this case. For this reason our work is complementary and deepens the understanding of the topics discussed in Ref. 30. In order to apply the above theory to JPEG images we should describe the related framework and the corresponding constraints. For large images of the order of \( 1024 \times 768 \) (in order to have statistical accuracy) a typical value of the size is \( 2.36 \) MB. Since we are doing double precision operations for noise estimation this size grows eight times to \( 18.88 \) MB. If we devote 1 GB of RAM then we can fit 50 images for noise estimation. For a specific coefficient, in order to estimate its variance, the 50 samples must lie with 90% probability in an interval of \( \pm 3\sigma \). For this reason if the quantization coefficient is \( Q \), in order to detect other values except the mean value the following inequality (Eq. (9)) must always hold.

\[
3\frac{\sigma}{Q} > 0.5 \Leftrightarrow \sigma > \frac{Q}{6}.
\] (9)

In order to find the allowed values of standard deviation we must find the possible values of the quantization coefficients. For JPEG quality 92, which is typical for cameras like Axis213PTZ, we use in our laboratories, we find the following quantization tables for the luminance (Table 1) and chrominance (Table 2).

**Table 1.** Luminance quantization table, \( Q = 92 \).

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>8</td>
<td>14</td>
<td>13</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>4</td>
<td>6</td>
<td>9</td>
<td>11</td>
<td>17</td>
<td>16</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>6</td>
<td>9</td>
<td>10</td>
<td>13</td>
<td>17</td>
<td>18</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>8</td>
<td>10</td>
<td>12</td>
<td>14</td>
<td>16</td>
<td>19</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>12</td>
<td>15</td>
<td>16</td>
<td>18</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1350003-11
We used the public domain software ImageMagick\textsuperscript{32} for compression of the images and the freeware JPEGSnoop\textsuperscript{33} program to obtain these tables. The existence of corner cases, in other words lower and upper bounds in the quantizer range, results in lack of Gaussian characteristics for values of mean smaller than $3\sigma$ and larger than $255 - 3\sigma$ in RGB color space. For this reason, if the above intensity values occupy a large portion of the image, our approach fails. We arbitrarily chose an upper bound of 12 for the standard deviation in order to minimize the reduction of the dynamic range. This value corresponds to a serious corruption of the first and last 36 values of the range.

### 3. Measurement Methodology

In this section, we record the experimental results that give quantitative answers to the above approximations. We used different contents in order to assure that our compensation procedure can provide effective results and extract sound conclusions. Providing results for different contents is crucial because different content provides different sites for stochastic linearization and different statistics of these sites. Though the JPEG procedure operates in practice on a $16 \times 16$ block at maximum (due to sub-sampling), having different contents leads to different realizations of these possible blocks. One could argue that we could generate randomly these blocks. Intuitively it is obvious that not all possible blocks correspond to real-life content. For this reason a selection of representative content could give real-world examples for performance evaluation. Large deviation in the results would show the necessity for more content. Fortunately we could quickly draw conclusions with a few representative examples as long as the variation is small. In the first set of measurements we record the mean square of the difference in the standard deviation between the linear distortion and the cases under consideration, namely the compensated and uncompensated distortions.
Our procedure is the following, we generated groups of 50 copies of a base uncompressed image (we use ASCII PPM file format). The images of each group were corrupted with additive noise of a group-specific value of variance. Different values of the variance correspond to different groups. We distinguish three cases. In the first case we applied on them only the linear part of the JPEG distortion. In other words we converted the image to YCbCr color space, we sub-sampled chroma components with the 2:1 scheme, we expanded it by duplication (no special up-sampling as it is usually the norm in the java.awt package or the open source libjpeg implementation which sets by default the value of doFancyUpsampling by default to true and was a source of various problems in our initial experiments) and converted back to RGB color space. For each pixel we estimated from the 50 corresponding values of the group members the mean and the variance and collected the results as a (mean, variance) pair. In the second case we included the nonlinear procedure which consists of the DCT transform, quantization and IDCT transform. We selected a quality of 92 which is common to the AXIS line of products used in our laboratory. In the third case we computed for each quantized DCT coefficient the mean and the standard deviation from the 50 corresponding values of the group members and applied our compensation procedure to these pairs. Subsequently, we renormalized the variance of the samples to the estimated variance and applied the IDCT with the following formula:

\[ c_{\text{renormalized}}(i) = \mu_{\text{input}} + \frac{\sigma_{\text{input}}}{\sigma_{\text{output}}}(c(i) - \mu_{\text{input}}). \]  

(10)

The flowchart of our compensation method is shown in Fig. 21. The idea behind this renormalization is to preserve correlation. We will come back to this point later in this chapter. We applied IDCT transform on the renormalized values to recover the blocks and expanded them with duplication. The rest of the procedure remained the same. Note the invariance of the mean image in both compensated and uncompensated distortion. From the mean value — standard deviation pairs per pixel we compute a curve that shows the dependence of the standard deviation on the pixel intensity. This is not a difficult task because for each interval between two consecutive integer values of the intensity we compute the mean of the standard deviation from pixels having mean value in this interval. In order to validate our approach we allowed a weak linear dependence on the noise variance and we neglect fixed pattern noise. Let us now come back to the explanation of the renormalization procedure. We could easily estimate the correlation coefficient between pairs of DCT coefficients in the same block and same layer of the color space (assuming a Gaussian distribution suffice to model the quantization effect). This way we could apply a linear transformation corresponding to IDCT and up-sampling to estimate inter-channel and spatial intra-channel correlation. If we selected this approach we had to make in the same block \(32 \times 64 = 2048\) pairwise estimations (the covariance matrix is symmetric). This would have resulted in cumbersome calculations. Instead we
employ the simpler procedure which is partially justified by the following reasoning. Suppose, we are given two Gaussian random variables \(X_1, X_2\). Their correlation matrix has the general form,

\[
C = \begin{pmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}.
\tag{11}
\]

Here \(\sigma_i\) is the standard deviation of Gaussian variable \(X_i\) and \(\rho\) is their correlation coefficient. Suppose we pass them through separate quantizers. Our objective is to find their correlation coefficient afterwards. In general the result should depend on the three previously mentioned numbers and the mean values of the coefficients. Figure 5 shows the results for various standard deviation pairs as a mean value over uniform samples of mean values.

Surprisingly the correlation coefficient exhibits great stability in its form for various values of the pair of input standard deviations. With good approximation we can say that it is insensitive to the quantization effect when members of the pair are not lying in the oscillatory region. For this reason, renormalization is expected to be an efficient solution with low complexity to the problem of correlation coefficient preservation after the compensation procedure. Of course our approach is not ideal but helps if we need correct estimation of the noise standard deviation before the quantizer. Alternatively we could use a look-up table solution which would be cumbersome due to the five-dimensional parameter space. There is another drawback more serious than the dimensionality of the look-up table. In order to have maximum accuracy each pair of variables must be compensated with a look-up table which is equivalent to the estimation of the 2048 covariance coefficients and subsequent compensation. For this reason, we resort to the more economical in term of calculations single coefficient compensation. In our third dataset we record the

![Fig. 5. Output correlation coefficient vs. input correlation coefficient for different standard deviation pairs.](image)
spatial intra-channel correlation. In order to present a coherent approach instead of using the difference in spatial correlation coefficient we use the following performance index called spread:

\[
\text{spread} = \sum |x\|y\| R(x, y)|. \quad (12)
\]

In the previous equation, \( R(x, y) \) is the normalized auto-correlation coefficient for a single channel. The meaning of the above index is that in a sense, it captures the spread of the autocorrelation. For this reason we use it to measure the similarity between the autocorrelation in the different cases. \( R(x, y) \) is estimated as a series of ergodic measurements for the frame sequence with length \( T \) at random sites of the image \((a_i, b_j)\), (after subtracting ergodic means and renormalizing with ergodic standard deviation per pixel) as

\[
R(x, y) = \frac{1}{N_aM_b} \sum_{i=1}^{N_a} \sum_{j=1}^{M_b} \frac{1}{T} \sum_{t=1}^{T} I(x + a_i, y + b_j, t)I(a_i, b_j, t). \quad (13)
\]

In order to have reliable results we allow values of the spatial variables \( x, y \) in the range \([-10, +10]\). We used 1000 random points per color. We must point out that for ideal white noise while \( T \to +\infty \) the quantity \( R \) should tend to zero. We show this behavior in the corresponding diagrams.

4. Simulation Results

We will compile three sets of measurements. We will try to uncover how our approach influences the precision in the recovery of noise standard deviation, inter-channel single pixel correlation and intra-channel spatial correlation. We believe that these measurements suffice to show the benefits and the limitations of our approach in camera noise estimation and they are standard practice in literature.

To this end we select a first dataset for the first series of measurements, which is shown in Fig. 6. As a first test of our approach we apply our theoretical results to a specific content (content 1) to find out whether the previous simplifications can offer an improvement over the existing practices. We select a noise model of the form,

\[
\begin{pmatrix}
\sigma_R^2 \\
\sigma_G^2 \\
\sigma_B^2
\end{pmatrix} = \begin{pmatrix}
12 + 0.01I_R \\
12 + 0.02I_G \\
12 + 0.03I_B
\end{pmatrix}. \quad (14)
\]

In the next three figures, Figs. 7–9, we show the measured variance in the different layers of the RGB color space. Some explanation in the figures is necessary. \( V \) is the value of the variance. \( V_{\text{uncomp}} \) is the variance estimated with the naive method. \( V_{\text{comp}} \) is the value of the variance estimated with our approach.

The above preliminary results show that our approach is better. Note also that both approaches are able to display correctly the linear dependence. In our tests the
slope was recovered with negligible error in all cases. However, we need quantitative results in order to measure the degree of success of compensation. For this reason we compute the mean square error between estimated and actual standard deviation separately for each color in the pixel intensity range 36–239 where we have shown to be region with approximately valid Gaussian properties. The results are plotted against different values of input noise standard deviation (noise variance in these tests has no dependence in pixel value) in the range 0.12 and for four different kinds of content. Results prefixed with $U_-$ are errors between uncompensated and linear distortion. Results prefixed with $C_-$ are errors between compensated and linear distortion. Figures 10–12 record the measured error for the three channels (red, green, blue) respectively.
Estimation of Frame Sequence Noise with Removal of JPEG Artifacts

Fig. 7. First tests of proposed approach for standard deviation estimation, red channel.

Fig. 8. First tests of proposed approach for standard deviation estimation, green channel.

We observe very small variations in the results between the four different contents which is an indication of a universal behavior. To this end we observe an improvement of 300% in the error performance between our approach and standard practice of just measuring the standard deviation of the pixel value. This should not come as a surprise since our approach is theoretically well justified. We observe that in the regime of standard deviation smaller than 4 our approach is not
better than the standard practice of just measuring the standard deviation of the pixel value. This can be attributed to two different factors. The first is the finite number of measurements. The more measurements we take, the better the Gaussian behavior of the system is displayed. The second is the fact that for small values of the input standard deviation the system is pushed towards the oscillatory region where our approach is approximate and generates errors. After this threshold the error in the usual method increases rapidly above 1.5 while our method converges rapidly to a constant small value near 0.4. A notable fact is that the error is bigger in the green color which could be attributed to the asymmetry of YCbCr transformation. One could expect better results while we refine the compensation lookup table. In our tests we observed marginal improvement when we decreased the step in $\sigma_{\text{input}}$ to values lower than 0.01. We also observed marginal improvement when we decreased the lookup table step for $\mu_{\text{output}}$ to values lower than 0.05. The lack of further improvement can be seen as a first limitation of our approach and can be attributed to lack of exact Gaussianity of the noise after the quantization procedure and the use of output mean value as an approximation to the input mean value to the look-up table.

The first dataset displays another limitation of our approach to the correct estimation of inter-channel correlation. Observing Figs. 13–15, the compensated and uncompensated distortion present big error in the calculation of the inter-channel correlation coefficient for values of the input noise standard deviation smaller than 5. This is to be expected for the uncompensated case because of the lack of compensation with the use of a look-up table.
Estimation of Frame Sequence Noise with Removal of JPEG Artifacts

Fig. 10. Error performance of our approach in standard deviation estimation for the first dataset, red channel.

However, despite the fact we compensate the standard deviation which leads to good accuracy in the estimation of the input standard deviation, the approach gives no real improvement in the estimation of inter-channel correlation. This should come as a surprise at first look. However, the correlation coefficient is an independent parameter which needs its own compensation techniques, something we did not account for in our calculations. As it is expected theoretically the estimation improves while the input standard deviation increases. The picture does not change in the second dataset (Fig. 16). For this reason we measure errors in the spatial intra-channel correlation spreading (Fig. 17). Here we observe that the proposed solution behaves slightly worse than the existing naïve approach. It becomes obvious that lack of correlation compensation leads to no further improvement than the classical approach. We can see again that the increase in input noise intensity improves the estimate at the expense of dynamic range. Here, we should point out that many cameras introduce complex noise reduction algorithms before the
JPEG compression step on the CCD sensor frame-buffer. The above results show clearly that the resulting noise profile cannot be recovered and modeled realistically.

Our approach shows promise in recovering partly the profile of input noise standard deviation, but with a big error. However, our approach and standard practice collapse when correlation is to be estimated. Only with big noise intensity it is possible to recover the underlying characteristics. For this reason, noise reduction should be done preferably after the storage step by using specially designed software. As a consequence, typical activation of image de-noising should be avoided if noise compensation and true camera noise modeling is desired.

In Fig. 18, we see the dependence of error on the compression quality of the image in both compensated and uncompensated case on content 3 with noise intensity 12. Both cases present a saturation effect while quality drops under 70. Our compensation case in this case has worst performance compared to the existing practice. However, in the range 85–100% quality our approach is not only smoother,
it is also much better in terms of estimation error. What is more than a mere coincidence is that this is the range of excellent visual quality with little artefacts (usually jaggies) and it seems that our measurements make this point stronger using an objective rather than a subjective criterion. After this point, we see a rapid deterioration in quality. It seems that this technique can be used to other image coders in order to assess their quality. An interesting question would be to measure the uncompensated error in the case of JPEG2000 in order to find a similar distinction in “high quality” and “low quality regions”. It is rather interesting the fact that the green color has the largest deviation which we again attribute to the asymmetry of YCbCr transform.

5. Real-World Measurements

For evaluating our approach towards the established practice of simple estimating the noise intensity as the variance of the pixel values we applied our theory to
V. I. Anagnostopoulos, E. S. Sardis & T. A. Varvarigou

Fig. 13. Estimated correlation coefficient between blue and red channels for the first dataset.

a real-camera. For this reason we do not know a priori the noise profile and for this reason we will use the available theory in order to find out what are the noise properties of the camera. We selected an AXIS 213PTZ camera\textsuperscript{36} since it is standard equipment in our laboratory. This camera comes with a mechanism to apply a de-noising algorithm to the images with three options, no de-noising, low de-noising and high de-noising. In the first case we selected to disable this mechanism. In the second case we selected the high de-noising option. We selected a place in the laboratory with good illumination conditions. After selecting a suitable shutter speed setting (in this case 1/150s) we grabbed 50 frames of the same scene (Fig. 19).

We tested the combination of de-noising and compensation. Consequently we made four measurements. In Fig. 20, we present our results on standard deviation and correlation coefficient. We have to note that the level of noise was unknown and we need to estimate it from the measurements. In the first diagram we show the measured variance in the red channel with the different combination of processes.
Estimation of Frame Sequence Noise with Removal of JPEG Artifacts

Fig. 14. Estimated correlation coefficient between green and blue channels for the first dataset.

We observe that the compensated values are the highest. We also observe that de-noising makes compensation more unstable. On average it is obvious that the difference between compensated and uncompensated standard deviation without noise reduction is 1. If we enable noise reduction this value becomes 1.5.

From the outlined theory we see that if we assume a constant noise level (which is suggested by the measurements) the compensated value must be within 0.5 from the real value. We also see that the uncompensated value must be within 0.8 from the real value. For this reason the real noise intensity must be approximately 4.5. In the case we enable the de-noising algorithm we see that because of error it is more difficult to estimate the real value of the noise. It is also striking the fact that this instability causes compensated reduced noise to have approximately the same value with uncompensated noise when reduction is disabled. We can make a rough estimate of 2.5 base on the mean error of the noise in case of denoising leading to 150% reduction due to the camera implemented algorithm. In order to complement our
measurements with more data showing the correctness of our approach we execute measurements on the inter-channel correlation coefficient between red and green channel. We see that compensation has a smoothing effect on the correlation coefficient due to closer resemblance to the original noise. In our case, we know that the error in RG varies between content but is bounded by 0.3 for noise intensity 4.5. We see again the instability caused by compensation. De-noising mathematically has no effect on the correlation coefficient if we assume white noise. This is observed in the case of uncompensated measurements but not in the case of compensated measurements. It is difficult to draw a safe conclusion as far as correlation is concerned. However, we must note that compensation gives a better approximation to a Gaussian process when the shape of the curves is the main concern and for this reason it seems preferable. For this reason it is not unreasonable to model the camera as a Gaussian noise source having nonzero correlation coefficients. The same could not be said a priori when compensation and de-noising are coupled due to the lack of theoretical evidence. A posteriori we see that this is not wrong.

Fig. 15. Estimated correlation coefficient between red and green channels for the first dataset.
Estimation of Frame Sequence Noise with Removal of JPEG Artifacts

Fig. 16. Second dataset for spatial spread comparison. Content 1 (upper left), content 2 (upper right), content 3 (lower left) and content 4 (lower right).

Fig. 17. Spatial spread comparison for uncompensated (left) and compensated noise (right) for the second dataset.
Fig. 18. Estimation error vs. JPEG compression quality for content 1 of the second dataset without compensation (left) and with compensation (right).

Fig. 19. Third dataset. Images of the lab from the AXIS 213 PTZ camera. Inherent de-noising de-activated (left) and de-noising activated (right).

Fig. 20. Standard deviation estimate for green channel (left) and correlation coefficient between red and green channel estimate (right).
6. Conclusions

In our tests it is clear that a theoretically well-justified approach can give significant improvement to the problem of noise estimation of camera noise with the use of compressed images. It is clear that we need big quality in the content to achieve significant accuracy in our estimation. Nevertheless we achieve a good estimation without using uncompressed images as it is currently proposed with the introduction of GiGE Vision cameras. In a system where bandwidth budget is of considerable importance and where we need on-line re-estimation of image noise, there are not many options and we present a better one under these constraints. For sufficiently big noise intensity corresponding to real-world cameras we show very good simulation results. Contrary to common practice, video noise reduction implemented in cameras must not be done prior to JPEG compression because it...
increases the distortion and alters significantly and artificially the noise properties of the image with respect to spatial intra-channel and inter-channel correlation. We also give quantitative bounds on the problem of compensation in the case of correlation and we show that it cannot be improved without significant computational overhead. In the case of high noise our compensation scheme gives good estimations to the aforementioned quantities and for this reason can capture better the underlying noise model. We also show that our methodology offers good estimation quality in the region of compression quality 85–100% and we show how similar tests can be executed on other well-known coders like JPEG2000.

Appendix

In our approach, we will use repeatedly the fact that a uniform quantizer under Gaussian noise approximately preserves mean value and variance over a wide range of values. We show this effect graphically. We record now the equations we used. It is not difficult to show that if random variable \( Y \) has a distribution \( f \) then using a uniform quantizer \( Q \) with integer values in 0.255 the following formula holds

\[
P\{Q(Y) \leq k\} = \int_{-\infty}^{\Phi(k + 0.5)} f(y)dy = P\{Y \leq \Phi(k + 0.5)\}, \quad (A.1)
\]

where the function \( \Phi \) is equal to

\[
\Phi(x) = \begin{cases} 
+\infty & x \in [255, +\infty) \\
 x & x \in (0, 255) \\
-\infty & x \in (-\infty, 0]
\end{cases}. \quad (A.2)
\]

The first and second moments can be computed readily by the equations

\[
E\{Q(Y)\} = \sum_{-\infty}^{+\infty} k \int_{\Phi(k - 0.5)}^{\Phi(k + 0.5)} f(y)dy, \quad (A.3)
\]

\[
E\{Q(Y)^2\} = \sum_{-\infty}^{+\infty} k^2 \int_{\Phi(k - 0.5)}^{\Phi(k + 0.5)} f(y)dy. \quad (A.4)
\]

From the above equation, it is not difficult to derive the well-known \( \sqrt{\frac{12}{\pi}} \) standard deviation formula for a uniform random variable in the quantizer’s range. We remark that function \( \Phi \) is equal to the identity function when the range of the quantizer is unbounded in both directions.

Acknowledgments

This work is supported by European Community’s Seventh Framework Programme FP7/2007-2013 under Grant Agreement No. 216465-SCOVIS.
References


1350003-29


**Vasileios I. Anagnostopoulos** received his Bachelor degree in Electrical and Computer Engineering from the National Technical University of Athens in 2000. He also received his M.Sc. in Applied Mathematics from the School of Mathematics and Applied Sciences of National Technical University of Athens in 2003. He is currently a Ph.D. candidate with the Department of Electrical and Computer Engineering from the National Technical University of Athens. He has been also involved in European Union projects, such as LION, OLYMPIC, POLYMNIA and SCOVIS.
Estimation of Frame Sequence Noise with Removal of JPEG Artifacts

Emmanuel S. Sardis received his Bachelor degree in Electrical and Computer Engineering from the National Technical University of Athens in 1996. He has also received Ph.D. degree from the same university in 2001 in the area of e-commerce and security. He is currently Visiting Assistant Professor in University of Aegean at the Department of Computer Engineering and Informatics and Senior Research Engineer in the Institute of Communication and Computer Systems (ICCS) and the National Technical University of Athens (NTUA). Dr. Sardis is Coordinator and Technical Manager of SCOVIS FP7 EU funded project. His research interests lie in the field of distributed and heterogeneous systems, artificial intelligence and multi-agent systems, workflows, and industrial environments using internet technologies. He has been also involved in numerous National and European Union projects, such as SocIoS, Hola!, SCOVIS, ANSWER, AGEDIS, ETIS-Agent, RePublic, TrainSEE, EUROSTAT, COMEXT, EUROFARM, ALEGRO, CORDIS, PROTONET.

Theodora A. Varvarigou received her B.Tech degree from the National Technical University of Athens, Athens, Greece in 1988, the M.Sc. degrees in Electrical Engineering (1989) and in Computer Science (1991) from Stanford University, Stanford, California in 1989 and the Ph.D. degree from Stanford University as well in 1991. She is now a Professor at the NTUA. Professor Varvarigou has great experience in the area of information processing, semantic web technologies, media streaming and casting and scheduling over distributed platforms. In this area, she has published more than 150 papers in leading journals and conferences. She has participated and co-ordinated several EU funded projects, related to the subject of the OSIA such as ANSWER, POLYMNIA, PENED-MPSE, GRIA, GRIDLAB, MEMPHIS, MKBEEB, MARIDES, DELOS, SCHEMA, FIDIS, HPC-EURORA, and so on.