Localization of Forgeries in MPEG-2 Video through GOP Size and DQ Analysis

D. Labartino #, T. Bianchi *, A. De Rosa #, M. Fontani †
D. Vázquez-Padín ‡, A. Piva #, M. Barni †
Outline

1. Motivation and Goal

2. Overall Structure of the Proposed System

3. Description of Two Main Steps
   - Localization of To-Be-Analyzed Frames
   - Double Quantization Analysis

4. Experimental Results
Video forensics today

- Many techniques for detecting double compression (probably, more than for images)
Video forensics today

• Many techniques for detecting double compression (probably, more than for images)

• Revealing video manipulation is much harder
  ▶ Naïve approach: double compression $\implies$ tampered video
  ▶ Not always the case (e.g. video may be re-encoded during device-to-computer transfer)

• Several works focusing on the removal/copying/replication of whole frames
Video forensics today

• Many techniques for detecting double compression (probably, more than for images)

• Revealing video manipulation is much harder
  ▶ Naïve approach: double compression $\Rightarrow$ tampered video
  ▶ Not always the case (e.g. video may be re-encoded during device-to-computer transfer)

• Several works focusing on the removal/copying/replication of whole frames

• Existing approaches for intra-frame forgery localization make strong assumptions, e.g. assuming only intra-coded frames, aka M-JPEG
Outline

1. Motivation and Goal
2. Overall Structure of the Proposed System
3. Description of Two Main Steps
   - Localization of To-Be-Analyzed Frames
   - Double Quantization Analysis
4. Experimental Results
Proposed Scheme

- Leveraging on the double compression undergone by forged video
Proposed Scheme

- Leveraging on the double compression undergone by forged video

- Use DQ analysis to localize the forged region within frames that were intra-coded twice

![Diagram showing the proposed scheme]

- Absence of Double Quantization traces

- Tampered video

- Tampered probability map

- Frame selection

- Tampering localization
Outline

1. Motivation and Goal

2. Overall Structure of the Proposed System

3. Description of Two Main Steps
   - Localization of To-Be-Analyzed Frames
   - Double Quantization Analysis

4. Experimental Results
MPEG-2 Coding 1/2

- **Frame types**
  - I: coded independently (much like JPEG)
  - P: predicted from a previous reference frame (I o P)
  - B: predicted from a previous and/or following reference frame (I o P)

- **Macroblob Types** (16x16 pixels)
  - intra-coded macroblocks (I-MB)
  - inter-coded macroblocks (P-MB)
  - skipped macroblocks (S-MB)
MPEG-2 Coding 2/2

- **DCT Coefficients Quantization**
  - Differently from JPEG, a quantization matrix is provided by the standard
  - Quantization strength adapted through a multiplier integer \( k \)
  - Different matrices are defined for I- and P- frames (we only care about I-frame quantization)

\[
Q_{i,j} = k \times \begin{pmatrix}
8 & 16 & 19 & 22 & 26 & 27 & 29 & 34 \\
16 & 16 & 22 & 24 & 27 & 29 & 34 & 37 \\
19 & 22 & 26 & 27 & 29 & 34 & 34 & 38 \\
22 & 22 & 26 & 27 & 29 & 34 & 37 & 40 \\
22 & 26 & 27 & 29 & 32 & 35 & 40 & 48 \\
26 & 27 & 29 & 32 & 35 & 40 & 48 & 58 \\
26 & 27 & 29 & 34 & 38 & 46 & 56 & 69 \\
27 & 29 & 35 & 38 & 46 & 56 & 69 & 83
\end{pmatrix}
\]
**DCT Coefficients Quantization**
- Differently from JPEG, a quantization matrix is provided by the standard
- Quantization strength adapted through a multiplier integer $k$
- Different matrices are defined for I- and P- frames (we only care about I-frame quantization)


**VBR coding**
- We assume Variable BitRate coding (VBR) $\Rightarrow k$ is fixed by the user
- Knowing $k$ gives all the quantization step $Q_{i,j}$ used in a compression
- This facilitates the task compared to JPEG
Variation of Prediction Footprint (VPF)\(^1\)

1. Description of Two Main Steps
2. Localization of To-Be-Analyzed Frames

---

D. Vázquez-Padín, M. Fontani, T. Bianchi, P. Comesaña, A. Piva, F. Pérez-González, M. Barni

*Detection of video double encoding with GOP size estimation*, WIFS 2012

---

Localization of Forgeries in MPEG-2 Video through GOP Size and DQ Analysis

---

D. Labartino et al.
Variation of Prediction Footprint (VPF)

First Compression

Second Compression

1 D. Vázquez-Padín, M. Fontani, T. Bianchi, P. Comesaña, A. Piva, F. Pérez-González, M. Barni

Detection of video double encoding with GOP size estimation, WIFS 2012
Variation of Prediction Footprint (VPF)¹

D. Vázquez-Padín, M. Fontani, T. Bianchi, P. Comesaña, A. Piva, F. Pérez- González, M. Barni
Detection of video double encoding with GOP size estimation, WIFS 2012

¹ Localization of Forgeries in MPEG-2 Video through GOP Size and DQ Analysis
Variation of Prediction Footprint (VPF)\(^1\)

- By estimating GOP\(_1\), we can locate those frames that have been intra-coded twice.

\(^1\) D. Vázquez-Padín, M. Fontani, T. Bianchi, P. Comesaña, A. Piva, F. Pérez-González, M. Barni

_Detection of video double encoding with GOP size estimation_, WIFS 2012
Variation of Prediction Footprint (VPF)\(^1\)

- By estimating \(GOP_1\), we can locate those frames that have been intra-coded twice.

\[\text{Diagram of I and P frames with highlighted frames indicating intra-coded frames.}\]

\(^1\) D. Vázquez-Padín, M. Fontani, T. Bianchi, P. Comesaña, A. Piva, F. Pérez- González, M. Barni
Detection of video double encoding with GOP size estimation, WIFS 2012
Variation of Prediction Footprint (VPF)\(^1\)

- By estimating GOP\(_1\), we can locate those frames that have been intra-coded twice.

---

\(^1\) D. Vázquez-Padín, M. Fontani, T. Bianchi, P. Comesaña, A. Piva, F. Pérez-González, M. Barni

*Detection of video double encoding with GOP size estimation, WIFS 2012*
Outline

1. Motivation and Goal
2. Overall Structure of the Proposed System
3. Description of Two Main Steps
   - Localization of To-Be-Analyzed Frames
   - Double Quantization Analysis
4. Experimental Results

Localization of Forgeries in MPEG-2 Video through GOP Size and DQ Analysis
D. Labartino et al.
From Double Quantization to Forgery Localization

First Compression → Local Tampering → Second Compression

- Single Quantization
- Double Quantization

---

1 Bianchi, De Rosa, Piva Improved DCT coefficient analysis for forgery localization in JPEG images, ICASSP’11
From Double Quantization to Forgery Localization

**First Compression**

1. $h_{DQ}(x)$
2. $h_{SQ}(x)$

**Local Tampering**

$\alpha h_{DQ}(x) + (1 - \alpha)h_{SQ}(x)$

**Second Compression**

$h_{mix}(x) = \alpha h_{DQ}(x) + (1 - \alpha)h_{SQ}(x)$
Estimation of $h_{SQ}$ (calibration technique)
Estimation of $h_{SQ}$ (calibration technique)

- $h_{cal}$ is an estimate of $h_0$ (histogram of unquantized coeffs)
- $h_{SQ}(x) \approx \tilde{h}(x) = \Delta_{k_2}(h_{cal})$
Estimation of $h_{DQ}$

- For a given $k_1$, $k_2$ and $Q$, we can count how many bins of $h_0$ fall in each bin of $h_{DQ}$

---

Bianchi, De Rosa, Piva. *Improved DCT coefficient analysis for forgery localization in JPEG images*, ICASSP’11
Estimation of $h_{DQ}$\(^1\)

- For a given $k_1$, $k_2$ and $Q$, we can count how many bins of $h_0$ fall in each bin of $h_{DQ}$
- Assuming $h_0$ locally uniform
  \[ h_{DQ}(x) \approx n(x; k_1, k_2, Q) \cdot \tilde{h}(x) \]

\(^1\)Bianchi, De Rosa, Piva. *Improved DCT coefficient analysis for forgery localization in JPEG images*, ICASSP’11
Estimation of $h_{DQ}$

- For a given $k_1$, $k_2$ and $Q$, we can count how many bins of $h_0$ fall in each bin of $h_{DQ}$
- Assuming $h_0$ locally uniform
  \[ h_{DQ}(x) \sim n(x; k_1, k_2, Q) \cdot \tilde{h}(x) \]

**MPEG-2 de-quantization formula differs from JPEG ones.**

The function $n(\cdot)$ was derived, resulting in:

\[
n(x; k_1, k_2, Q) = \frac{k_1 \times Q}{16} \left( \left\lfloor \frac{16}{Q \times k_1} \left\lfloor \frac{k_2 \times Q}{16} \left( x + \frac{1}{2} \right) \right\rfloor \right\rfloor - \left\lfloor \frac{16}{Q \times k_1} \left\lfloor \frac{k_2 \times Q}{16} \left( x - \frac{1}{2} \right) \right\rfloor \right\rfloor \right)
\]

---

1 Bianchi, De Rosa, Piva. *Improved DCT coefficient analysis for forgery localization in JPEG images*, ICASSP’11
Description of Two Main Steps

Double Quantization Analysis

**Estimation of** $h_{DQ}$ ¹

- For a given $k_1$, $k_2$ and $Q$, we can count how many bins of $h_0$ fall in each bin of $h_{DQ}$
- Assuming $h_0$ locally uniform
  
  
  $h_{DQ}(x) \approx n(x; k_1, k_2, Q) \cdot \tilde{h}(x)$

**MPEG-2 de-quantization formula differs from JPEG ones.**

*The function $n(\cdot)$ was derived, resulting in:*

$$n(x; k_1, k_2, Q) = \left\lfloor \frac{k_1 \times Q}{16} \left( \left\lfloor \frac{k_2 \times Q}{16} \left( x + \frac{1}{2} \right) \right\rfloor \right\rfloor - \left\lfloor \frac{k_1 \times Q}{16} \left( x - \frac{1}{2} \right) \right\rfloor \right\rfloor$$

---

¹ Bianchi, De Rosa, Piva *Improved DCT coefficient analysis for forgery localization in JPEG images*, ICASSP’11
Estimating $k_1$

Model: 

$$\hat{h}_{mix}(x; k_1, \alpha) = \alpha \cdot n(x; k_1) \cdot \hat{h}(x) + (1 - \alpha) \cdot \hat{h}(x)$$

$h_{DQ}$ estimate

$h_{SQ}$ estimate
Estimating $k_1$

Model

\[ \tilde{h}_{mix}(x; k_1, \alpha) = \alpha \cdot n(x; k_1) \cdot \tilde{h}(x) + (1 - \alpha) \cdot \tilde{h}(x) \]

- $h_{DQ}$ estimate
- $h_{SQ}$ estimate

Error

\[ e(k_1, \alpha) = \sum_{x \neq 0} \left[ h_{mix}(x) - \tilde{h}_{mix}(x; k_1, \alpha) \right]^2 \]
Estimating $k_1$

Model

$$\tilde{h}_{mix}(x; k_1, \alpha) = \alpha \cdot n(x; k_1) \cdot \tilde{h}(x) + (1 - \alpha) \cdot \tilde{h}(x)$$

Error

$$e(k_1, \alpha) = \sum_{x \neq 0} \left[ h_{mix}(x) - \tilde{h}_{mix}(x; k_1, \alpha) \right]^2$$

Estimate $k_1$ and $\alpha$ by minimizing $e(k_1, \alpha)$
Probability map

Knowing $k_1$ allows us to write, for the $i$-th coefficient $x_i$:

$$p(x_i | T) = \tilde{h}(x_i)$$ $$p(x_i | O) = n(x_i; k_1) \cdot \tilde{h}(x_i).$$
Knowing \( k_1 \) allows us to write, for the \( i \)-th coefficient \( x_i \):

\[
p(x_i|T) = \tilde{h}(x_i) \quad p(x_i|O) = n(x_i; k_1) \cdot \tilde{h}(x_i).
\]

By Bayes rule, and assuming equal priors:

\[
p(T|x_i) = \frac{P(x_i|T) \cdot P(T)}{P(x_i|T) \cdot P(T) + P(x_i|O) \cdot P(O)} = \frac{1}{1 + n(x_i, k_1)}
\]
Probability map

Knowing \( k_1 \) allows us to write, for the \( i \)-th coefficient \( x_i \):

\[
p(x_i|T) = \tilde{h}(x_i) \quad p(x_i|O) = n(x_i; k_1) \cdot \tilde{h}(x_i).
\]

By Bayes rule, and assuming equal priors:

\[
p(T|x_i) = \frac{P(x_i|T) \cdot P(T)}{P(x_i|T) \cdot P(T) + P(x_i|O) \cdot P(O)} = \frac{1}{1 + n(x_i, k_1)}
\]

**HYPOTHESIS**

Statistical independence between coeffs within a block

\[
P_B = \frac{1}{\prod_{i|x(i) \neq 0} n(x_i; k_1) + 1}
\]
Probability map

Knowing $k_1$ allows us to write, for the $i$-th coefficient $x_i$:

$$p(x_i|T) = \tilde{h}(x_i) \quad p(x_i|O) = n(x_i; k_1) \cdot \tilde{h}(x_i).$$

By Bayes rule, and assuming equal priors:

$$p(T|x_i) = \frac{P(x_i|T) \cdot P(T)}{P(x_i|T) \cdot P(T) + P(x_i|O) \cdot P(O)} = \frac{1}{1 + n(x_i, k_1)}$$
Outline

1. Motivation and Goal

2. Overall Structure of the Proposed System

3. Description of Two Main Steps
   - Localization of To-Be-Analyzed Frames
   - Double Quantization Analysis

4. Experimental Results
Dataset

Dataset built from 7 “raw” videos available in the Internet (resolution: 720x576 pixels)

- ducks_take_off.y4m
- in_to_tree.y4m
- old_town_cross.y4m
- park_joy.y4m
- shields.y4m
- sunflower.y4m
- touchdown_pass.y4m
Chosen parameters for creating the dataset (112 video):

- $\text{GOP}_1 = 12, \text{ GOP}_2 = 15$
- $k_1 \in \{12, 16, 20, 24\}$
- $k_2 \in \{4, 6, 8, 10\}$
Chosen parameters for creating the dataset (112 video)

- $GOP_1 = 12, \quad GOP_2 = 15$
- $k_1 \in \{12, 16, 20, 24\}$
- $k_2 \in \{4, 6, 8, 10\}$
Localization example

Tampered frame

Ground truth

Probability map
Experimental Results

ROC curves

$k_1 = 12$

$k_1 = 16$

$k_1 = 20$

$k_1 = 24$
## Experimental Results

**ROC curves**

AUC obtained with the proposed method:

<table>
<thead>
<tr>
<th>$k_1$</th>
<th>$k_2$</th>
<th>4</th>
<th>6</th>
<th>8</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
<td>0.98</td>
<td>0.97</td>
<td>0.68</td>
<td>0.63</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>0.98</td>
<td>0.98</td>
<td>0.96</td>
<td>0.93</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>0.98</td>
<td>0.98</td>
<td>0.91</td>
<td>0.94</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>0.98</td>
<td>0.98</td>
<td>0.97</td>
<td>0.94</td>
</tr>
</tbody>
</table>
Future works

Relax constraints:
Future works

Relax constraints:

- Robustness to frame de-synchronization
- Variable GOP
- Presence of B-frames
Future works

Relax constraints:

- Robustness to frame de-synchronization
- Variable GOP
- Presence of B-frames

\[ k_2 > k_1 \]

\[ \text{VPF} \]

\[ \downarrow \]

Localization of Forgeries in MPEG-2 Video through GOP Size and DQ Analysis

D. Labertino et al.
Future works

Relax constraints:

- Robustness to frame de-synchronization
- Variable GOP
- Presence of B-frames

- Varying $k$ throughout the frame (CBR coding mode)
  - $k_2 > k_1$
  - Cropping detection

⇓

VPF

Localization of Forgeries in MPEG-2 Video through GOP Size and DQ Analysis
Future works

Relax constraints:

- Robustness to frame de-synchronization
- Variable GOP
- Presence of B-frames

⇓

VPF

- Varying $k$ throughout the frame (CBR coding mode)
- $k_2 > k_1$
- Cropping detection

⇓

Localization Algorithm
Thanks for your attention.

This work was partially supported by the REWIND project.
More info at www.rewindproject.eu