Towards copy-evident JPEG images

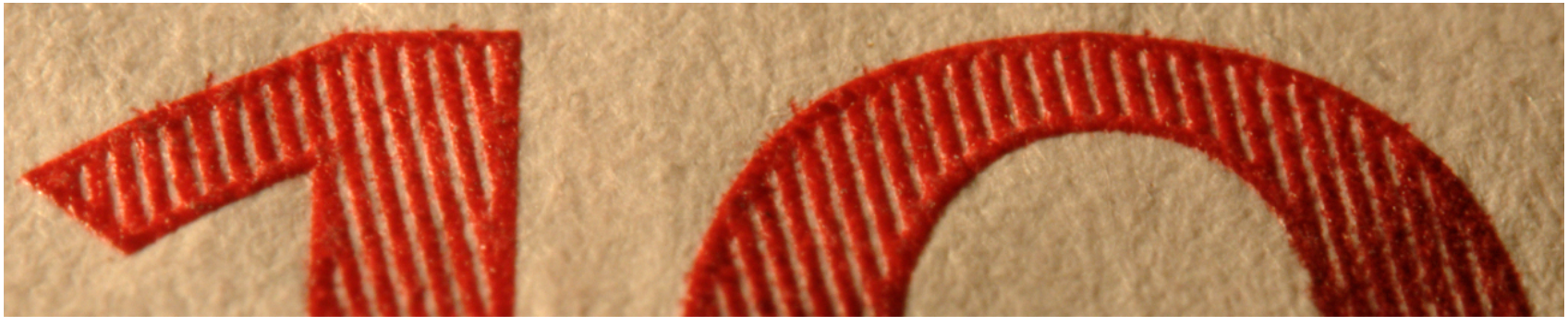
Andrew B. Lewis and Markus G. Kuhn

Computer Laboratory

Informatik 2009: Workshop Digitale Multimedia-Forensik –
Techniken und Anwendungsgebiete

Physical document security

- Documents of value (currency, etc.) may use anti-counterfeiting security features
- Expensive to produce an identical copy
- Use special materials (e.g. metallic strips), intaglio printing, offset printing, chemicals, holograms, kinegrams, ...
- Naïve duplication may reveal a hidden message, or simply cause visible artifacts to appear which de-value the document



Security printing (1)

- Most counterfeiters try to use consumer equipment: digital scanning and printing
- Hidden information is modulated onto a printable carrier, consisting of screen elements (dots, lines, . . .).

Examples:

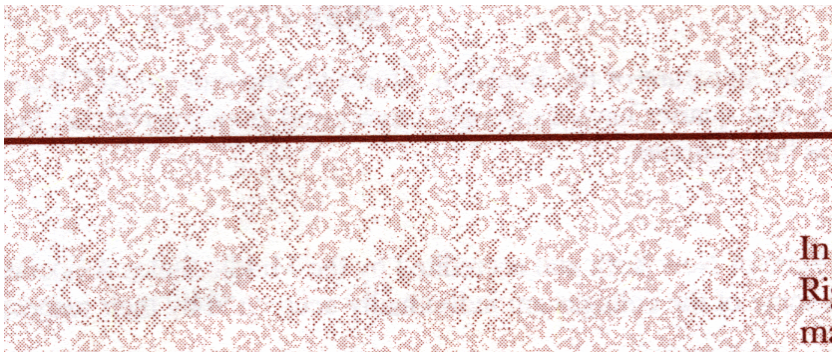
- Screen angle modulation
- Line frequency trap
- Frequency modulation of minimal dots
  - Defeats anti-aliasing filter scan-trap countermeasure

original note

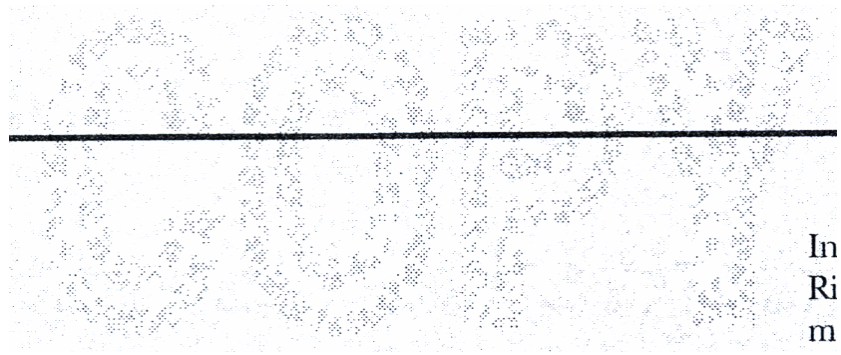
digital scan

Security printing (2)

- Concentric screens (moiré), dot shape modulation, . . . [1]
- When the spatial frequency of carrier patterns is sufficiently high, the naked eye cannot resolve the carrier screen and a uniform field is observed.



original document



photocopy

[1] Rudolf L. van Renesse *Hidden and scrambled images – a review* in *Proceedings of SPIE*, volume 6477, page 333, 2002.

Copy evidence in digital media

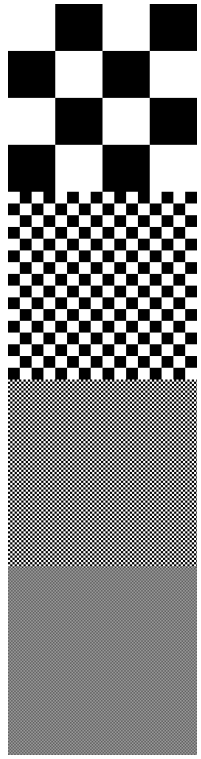
- Are similar techniques possible with digital formats?
- Can we add imperceptible patterns to an original image, video or audio signal that are perceptible after copying?
- Copying means standard lossy signal processing, such as recompression and resampling.

Applications:

- Protect valuable content which might be distributed to content sharing website
- Visible warning when quality has been degraded by a hidden processing step

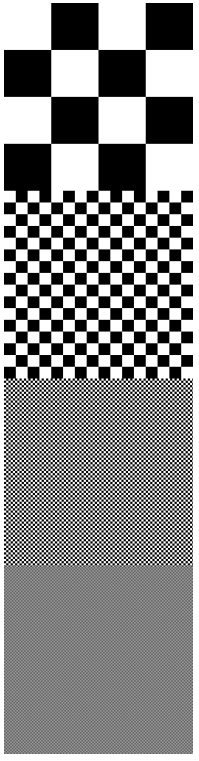
Possible techniques

- Regions of a single high spatial frequency are perceived as uniform
- Low frequency differences are more noticeable than high frequency differences
- Artifacts of lossy processing that could be exploited to uncover a message:
  - Non-linearities:
    gamma correction, quantization, clipping
  - Artifacts:
    aliasing, blocking



Possible techniques

- Regions of a single high spatial frequency are perceived as uniform
- Low frequency differences are more noticeable than high frequency differences
- Artifacts of lossy processing that could be exploited to uncover a message:
  - Non-linearities:
    gamma correction, quantization, clipping
  - Artifacts:
    aliasing, blocking

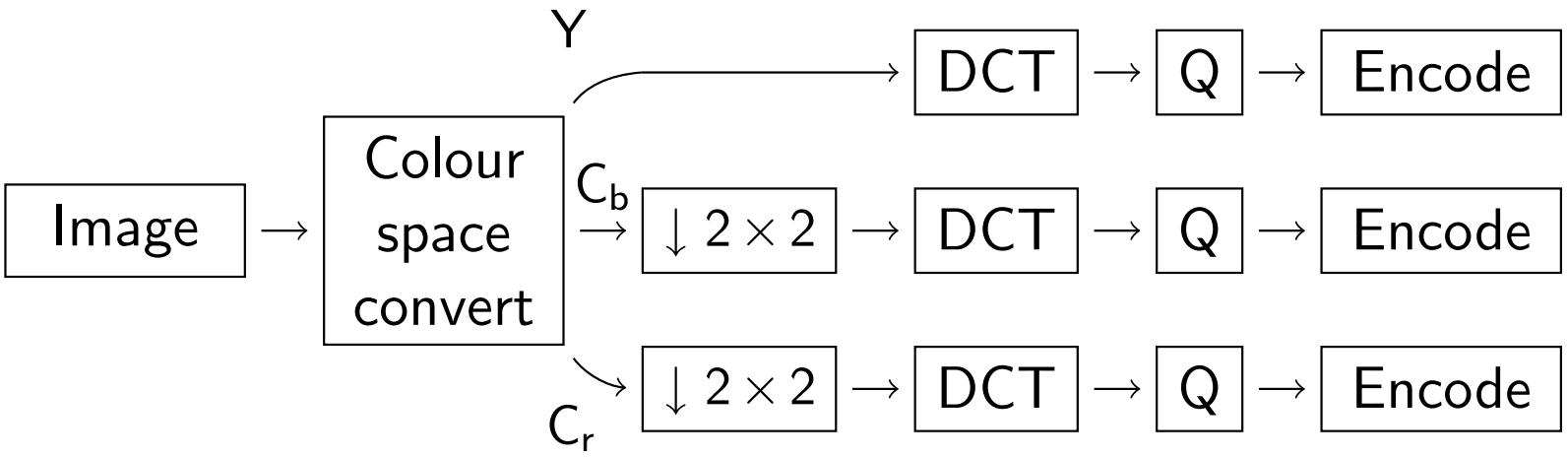


Approach

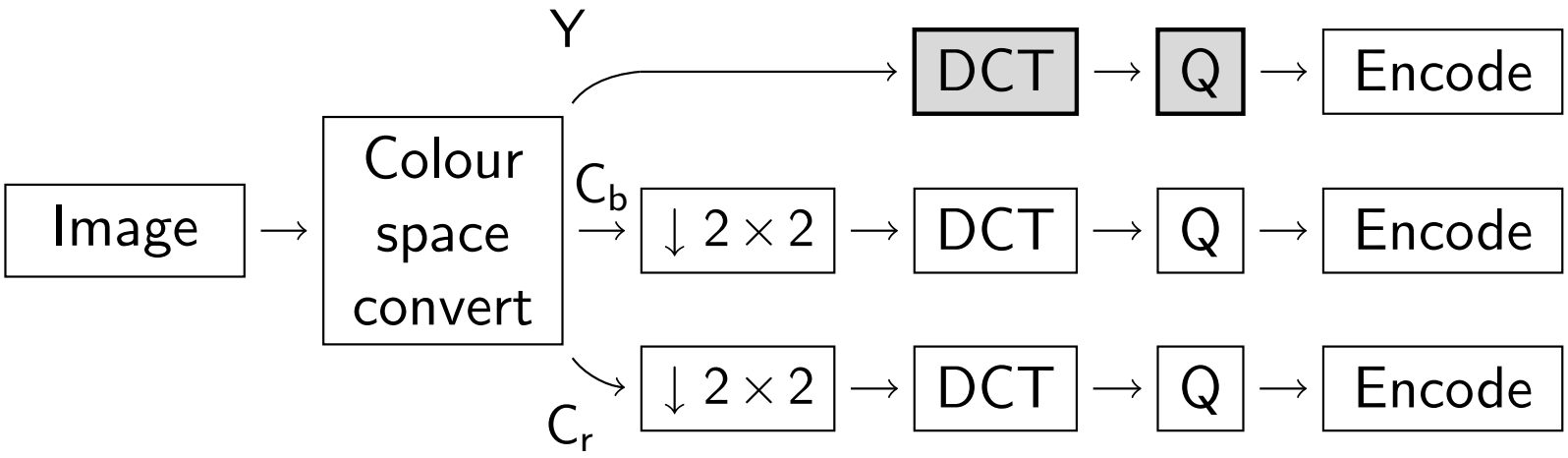
Difficult problem: compression algorithms try to *minimize* perceptible distortion

- Know the compressor, so can select worst case
- Write bitstream directly to give precise control over values
- Targeted or untargeted: known recompression parameters?
- This paper: initial exploration
  - JPEG recompression
  - Known quantization matrix
  - Uniform image region

Outline of the JPEG algorithm



Outline of the JPEG algorithm



Discrete cosine transform

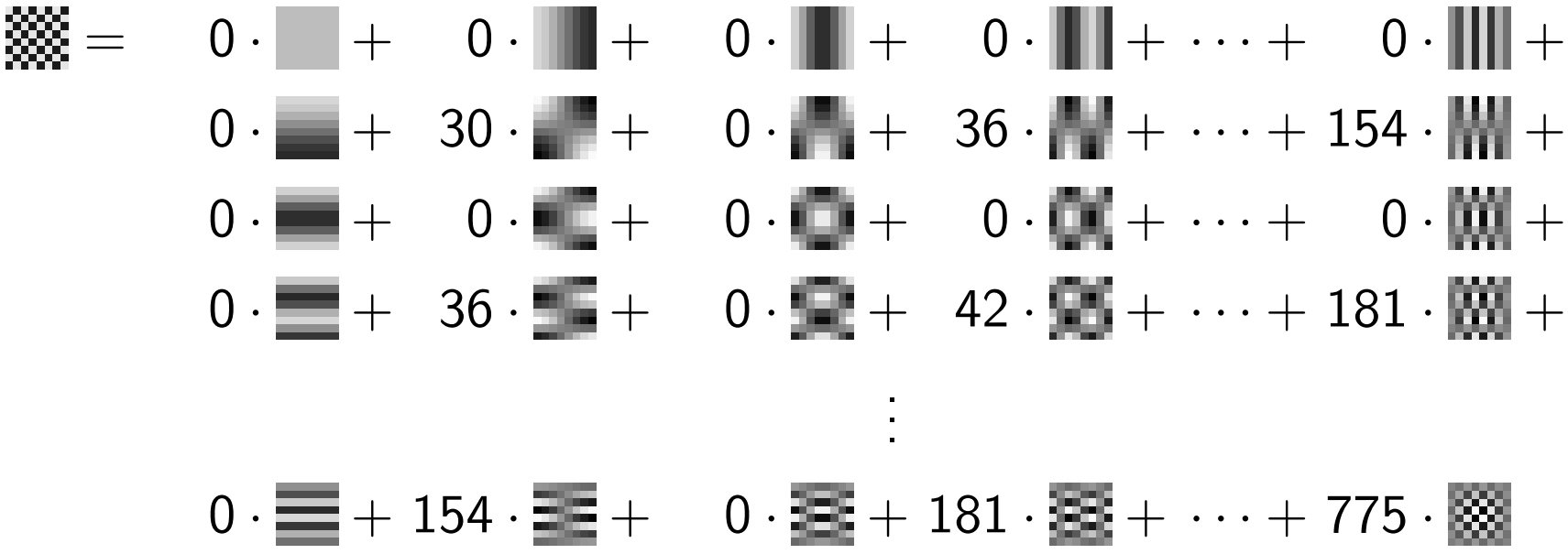
DCT decomposes 8×8 block of samples $s_{i,j}$ into weighted sum:

$$
\begin{aligned}
s = {} & S_{0,0} \cdot \text{  } + S_{0,1} \cdot \text{  } + S_{0,2} \cdot \text{  } + S_{0,3} \cdot \text{  } + \cdots + S_{0,7} \cdot \text{  } + \\
& S_{1,0} \cdot \text{  } + S_{1,1} \cdot \text{  } + S_{1,2} \cdot \text{  } + S_{1,3} \cdot \text{  } + \cdots + S_{1,7} \cdot \text{  } + \\
& S_{2,0} \cdot \text{  } + S_{2,1} \cdot \text{  } + S_{2,2} \cdot \text{  } + S_{2,3} \cdot \text{  } + \cdots + S_{2,7} \cdot \text{  } + \\
& S_{3,0} \cdot \text{  } + S_{3,1} \cdot \text{  } + S_{3,2} \cdot \text{  } + S_{3,3} \cdot \text{  } + \cdots + S_{3,7} \cdot \text{  } + \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\
& S_{7,0} \cdot \text{  } + S_{7,1} \cdot \text{  } + S_{7,2} \cdot \text{  } + S_{7,3} \cdot \text{  } + \cdots + S_{7,7} \cdot \text{  }
\end{aligned}
$$

Weights $S_{i,j}$ are DCT coefficients.

Discrete cosine transform

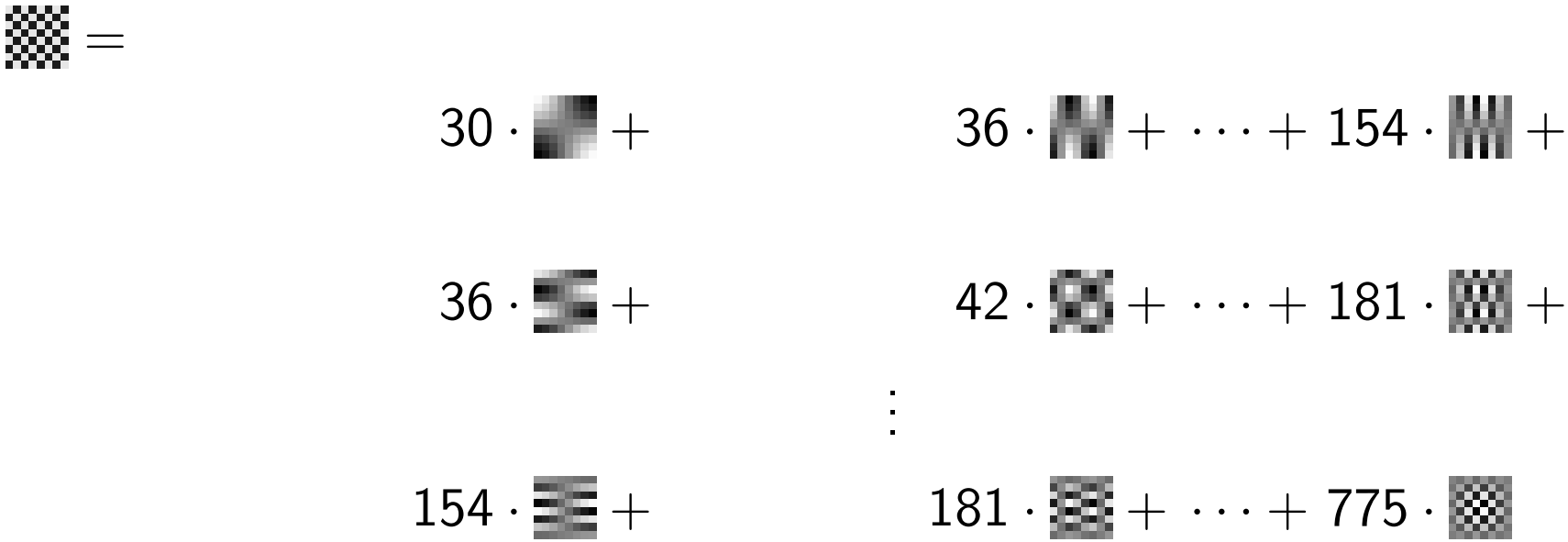
DCT decomposes 8×8 block of samples $s_{i,j}$ into weighted sum:



Weights $S_{i,j}$ are DCT coefficients.

Discrete cosine transform

DCT decomposes 8×8 block of samples $s_{i,j}$ into weighted sum:



Weights $S_{i,j}$ are DCT coefficients.

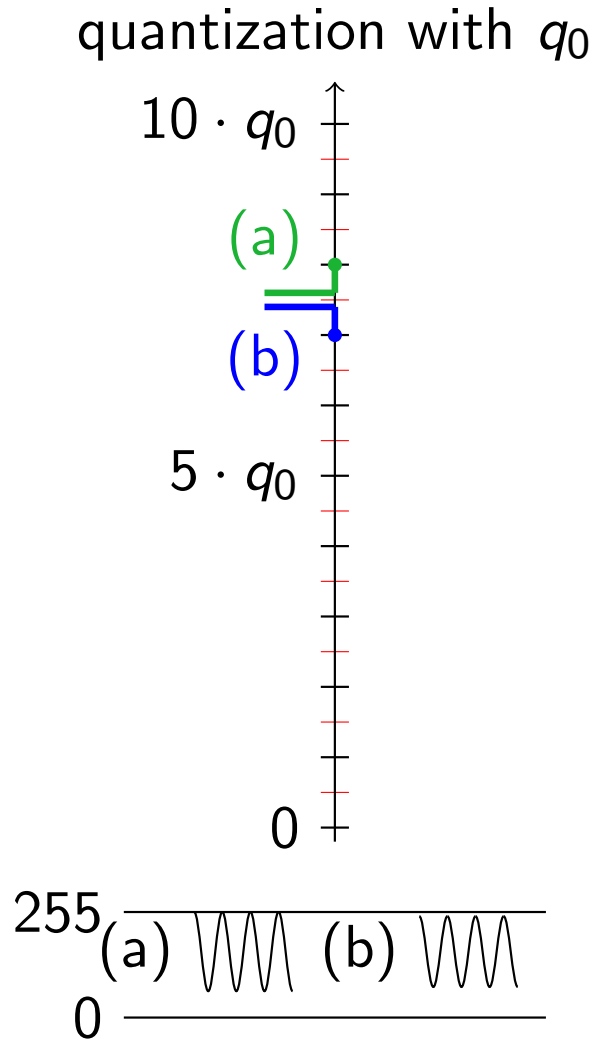
Quantization

- Quantization:

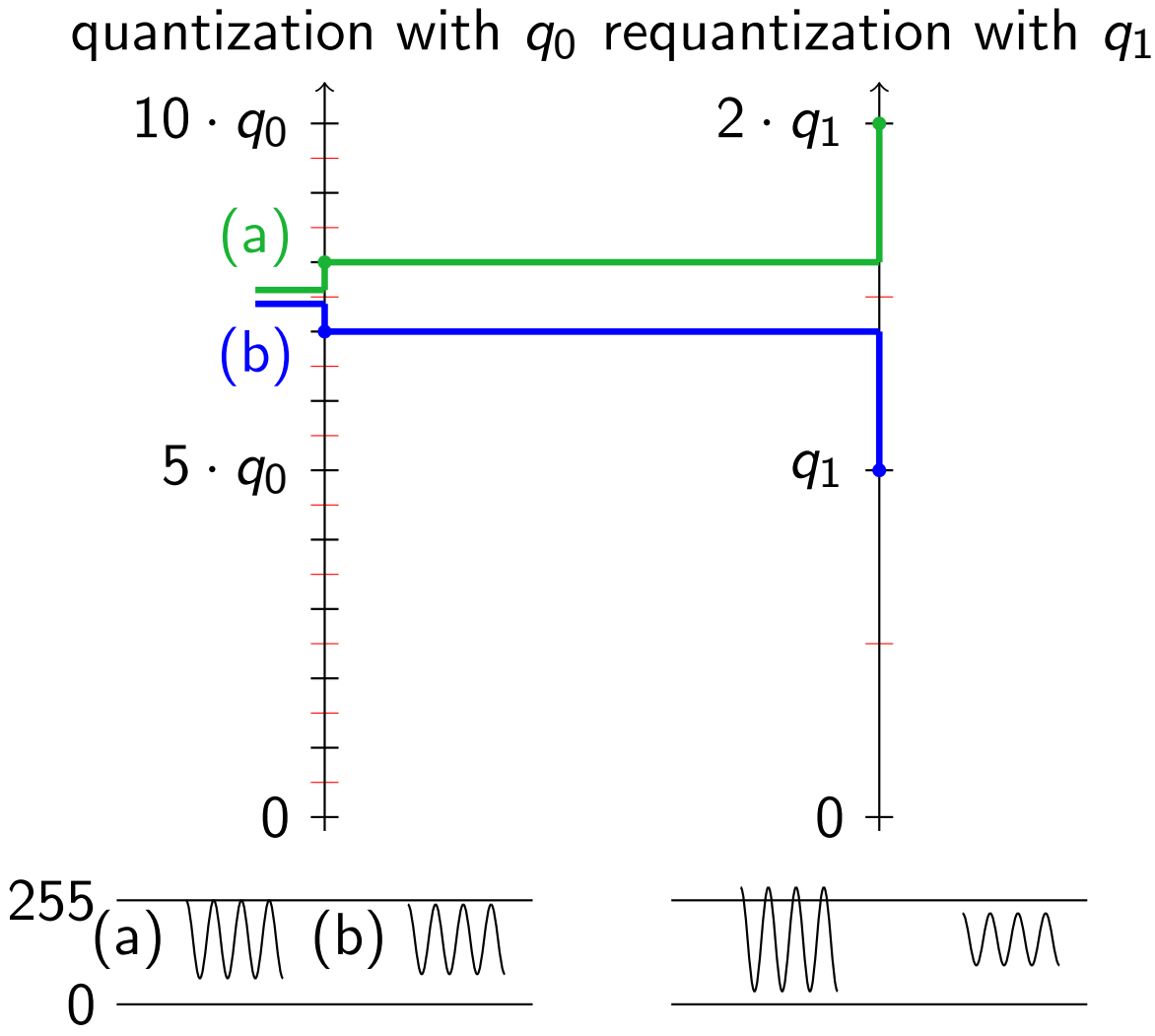
$$\hat{X}_{i,j} = \operatorname{sgn}(X_{i,j}) \cdot \left\lfloor \frac{|X_{i,j}| + \lfloor Q_{i,j}/2 \rfloor}{Q_{i,j}} \right\rfloor$$

- Dequantization

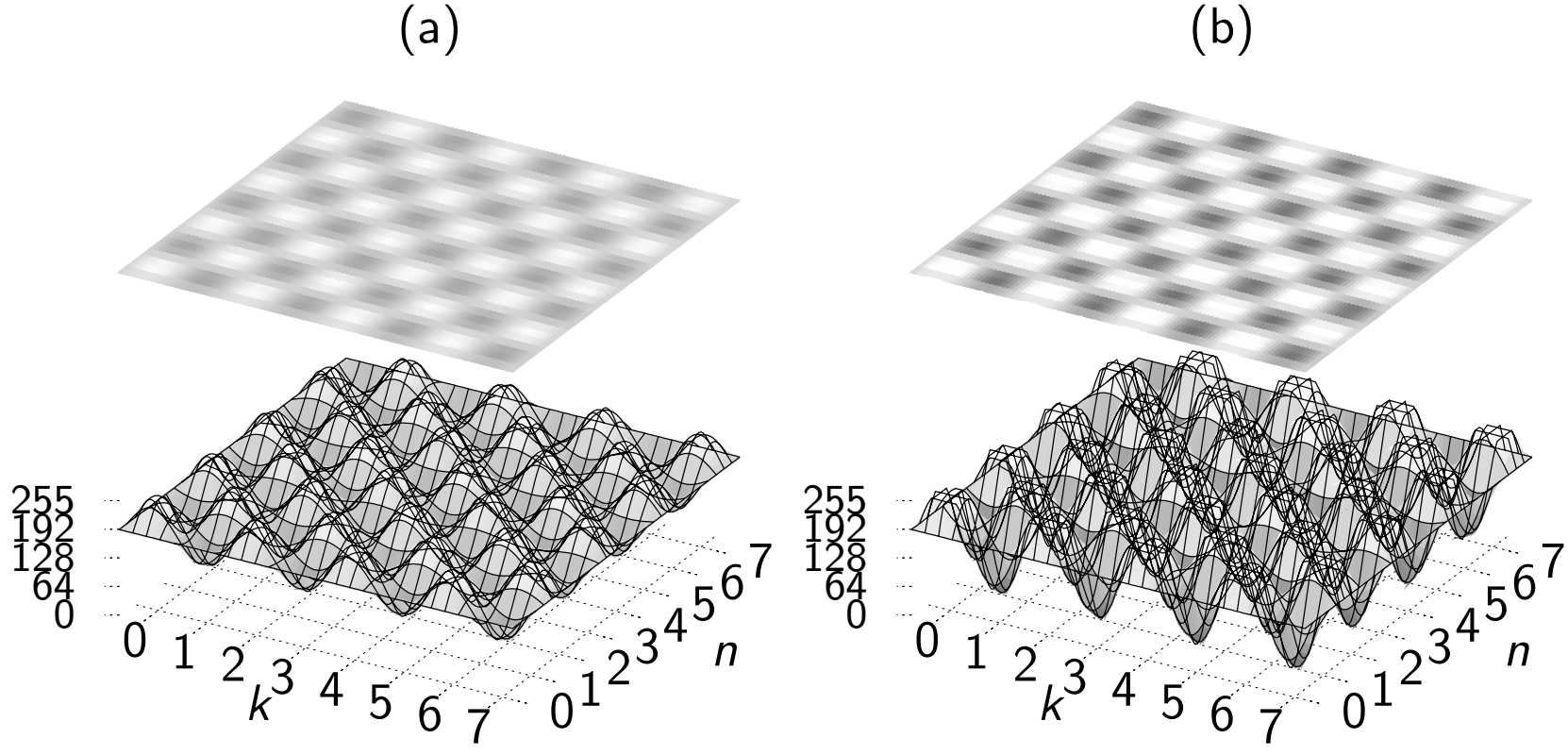
$$X'_{i,j} = Q_{i,j} \cdot \hat{X}_{i,j}$$



Requantization



Clipping after requantization



Marking algorithm

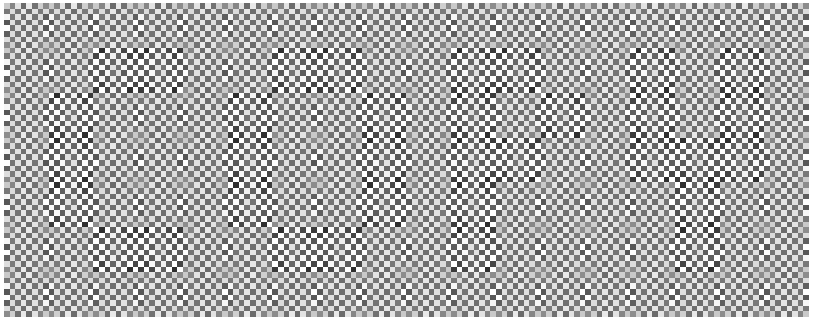
- Each bi-level message pixel maps to one 8×8 DCT block
- Add checkerboard pattern to block
- Amplitude of pattern chosen so that:
  - Foreground message blocks use closest higher amplitude above some quantization decision boundary
  - Background message blocks use closest lower amplitude below some quantization decision boundary
  - Clipping occurs after IDCT in recompressed image foreground blocks
- In the recompressed image, foreground message blocks appear darker than background message blocks
- In the marked image, foreground and background blocks appear the same

Example

The message to be embedded:

A uniform grey image is replaced with a checkerboard pattern with the same perceived brightness:

The result of recompression with a particular lower quality factor:



Summary

- We have demonstrated a copy-evident multimedia file, in which a human-readable message becomes visible after recompressing the original image.
- Our algorithm is applicable to uniform regions in images which will be recompressed with specific quantization settings.

Further work:

- Extend the marking process to handle arbitrary photographs
- Untargeted mark for JPEG images, not tied to particular recompression quantization matrix
- Audio and video signals