A Study of Embedding Operations and Locations for Steganography in H.264 Video

Andreas Neufeld and Andrew D. Ker

Oxford University Department of Computer Science

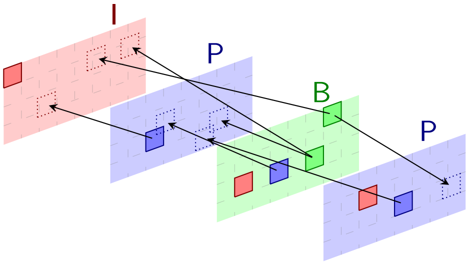
Andreas Neufeld is now with the Image and Pattern Analysis Group, University of Heidelberg, Germany.

February 6, 2013

Video Compression

- Similar to JPEG, H.264 uses an DCT approximation to code video data. DCT block size is 4×4.
- In H.264, residuals are transformed using the approximate DCT, these are image data minus a prediction.
- Chroma DC coefficients of a 16×16 macroblock are compressed with an additional Hadamard transform.
- There are three frame types: I, P and B frames.

The Different Frame Types



- I frames do not depend on other frames.
- P and B frames use *inter-prediction*, they refer to other frame(s) for prediction.
- P frames use a single past frame, B frames use two past or future frames.

Outline

Questions About Embedding

1. Which are the best embedding operations for steganography in H.264?
   - Which of LSBR, LSBM and F5 is least detectable?
   - How detectable are embeddings in coefficients of small magnitude?
2. Which are the best embedding locations?
   - P slices only, or both P and B slices?
   - Luma and/or Chroma channels?
   - Shall we use the DC coefficients for embedding?
3. Which of low quality and high quality videos is better suited for embedding?

Experiments

- We use a set of 16 DVDs, transcoded with x264 in two quality settings:
  - low quality (500kbit/s)
  - high quality (3000kbit/s)
- Embedding is *simulated* in the decoder.
- Features are extracted and stored for the entire collection.
- We test 7 different operations and 12 different locations, for two quality settings.
- This makes 168 feature collections in total.

Features

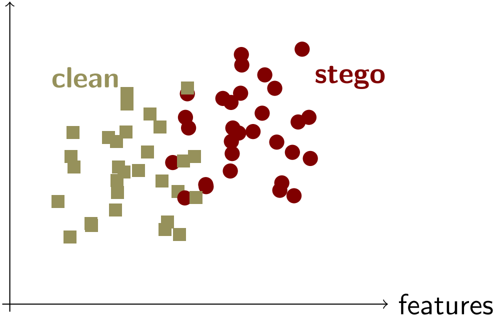
Our featureset consists of three parts:

Histogram	The histograms for each coefficient. Ranges differ, going down to zero for very rare coefficients. Dimension is 269.
Co-occurrence	The histogram for each coefficient pair. Dimension is 1015.
$U \times V$	The histogram of coefficient pairs between the two Chroma channels (U and V). Dimension is 598.

- Total dimension is 1882.
- Our sets consist of approximately 20,000 feature vectors.

The Information-Theoretic Model

We embed in a large set of videos, then measure separability of clean and stego feature vectors.



The Kullback-Leibler Divergence (KL-D)

The KL-D is an information-theoretic distance measure on probability distributions:

$$D_{KL}(p_c, p_s) = \sum_x p_c(x) \log \frac{p_c(x)}{p_s(x)}$$

- The KL-D relates directly to the performance of an optimal detector.
- Higher KL-D $\rightarrow$ more detectable.
- It is difficult to estimate.
- Using Gaussian distributions fails because of degenericity.

The Maximum Mean Discrepancy (MMD)

$$MMD(\mathcal{F}, p, q) = \sup_{f \in \mathcal{F}} \left(\mathbb{E}_{x \sim p} f(x) - \mathbb{E}_{x \sim q} f(x) \right)$$

Where $\mathcal{F}$ is a suitable function class and p, q are probability distributions.

- tells us how well a certain function class can separate two distributions.
- In our case, this function class corresponds to Gaussian kernels, which have been used successfully in SVMs.

MMD Estimator

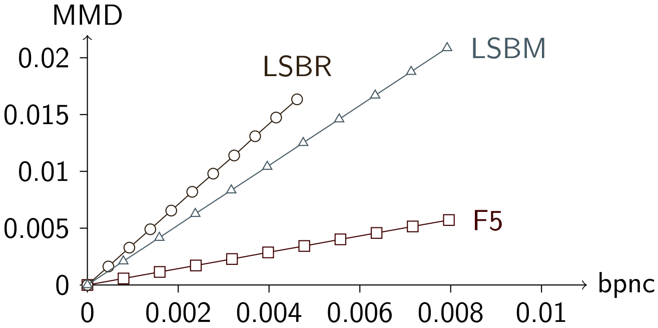
$$MMD(X,Y) \approx \sqrt{\frac{1}{N(N-1)} \sum_{i \neq j} k(x_i, x_j) - 2k(x_i, y_j) + k(y_i, y_j)}$$

With a Gaussian kernel:

$$k(x,y) = \exp(-\gamma \|x-y\|^2)$$

We set $\gamma = \eta^{-2}$ where η is the median of L_2 distances of all pairs of clean feature vectors.

The Maximum Mean Discrepancy (MMD)



- Embedding was done in Luma only.
- Different ranges on the x-axis indicate the use of different coefficients.

Questions About Embedding

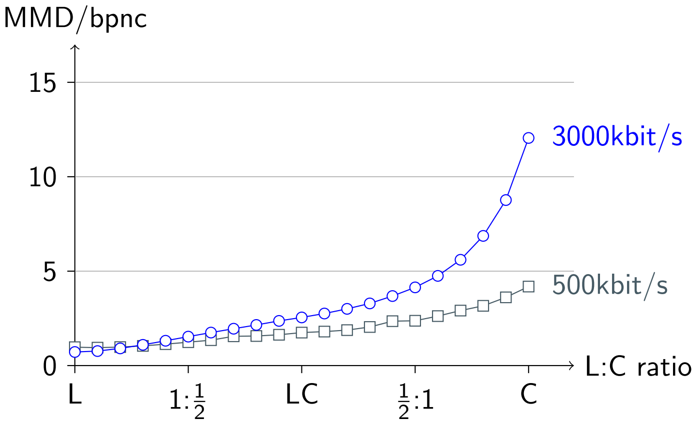
1. Which are the best embedding operations for steganography in H.264?
   - Which of LSBR, LSBM and F5 is least detectable?
   - How detectable are embeddings in coefficients of small magnitude?
2. Which are the best embedding locations?
   - P slices only, or both P and B slices?
   - Luma and/or Chroma channels?
   - Shall we use the DC coefficients for embedding?
3. Which of low quality and high quality videos is better suited for embedding?

The Maximum Mean Discrepancy (MMD)

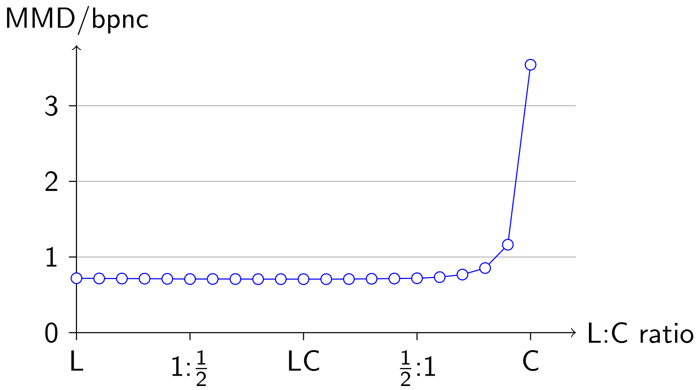
- It has been proven that the MMD is locally linear near zero.
- We use MMD/bpnc slope for comparison.
- We have 168 test sets, each containing 10 featuresets of size $\approx 20,000$ vectors with dimension 1882 $\rightarrow \approx$ 471GB of binary feature data.

	AC+DC			AC		
	Luma	Chroma	Both	Luma	Chroma	Both
F5	**0.72**	12.05	2.55	1.18	3.62	1.07
LSBM	2.63	6.84	2.50	4.49	8.80	4.22
LSBR	3.54	9.61	3.42	6.20	12.06	5.84

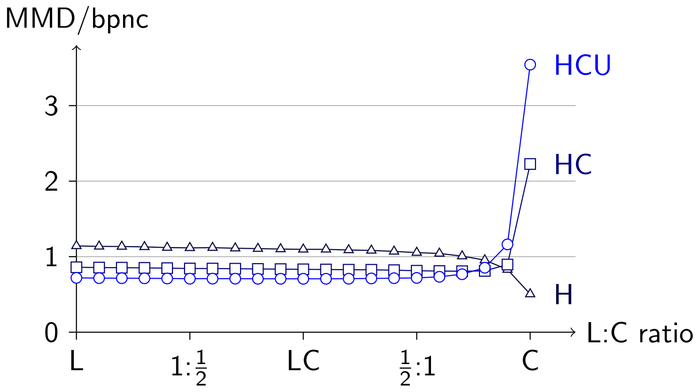
Balancing payload between Luma and Chroma



Using all Luma and Chroma AC



Varying the Feature Parts



Our Answers

1. - As in images, F5 is the least detectable embedding method.
   - We have to exclude zero coefficients. Excluding small magnitude coefficients increases detectability.
2. - B slices increase detectability.
   - We achieve lowest detectability with using F5 in Luma and Chroma AC.
   - In many other cases the Chroma DC coefficients decrease detectability.
3. High quality videos yield a lower detectability.

Thank You!