

Set-Membership Identification of Resampled Signals

David Vázquez-Padín¹, Pedro Comesaña¹, and
Fernando Pérez-González^{1,2}

1. Signal Theory and Communications Department, University of Vigo, Vigo, Spain.
2. Gradiant (Galician Research and Development Center in Advanced Telecommunications),
Vigo, Spain.

Guangzhou, China. November 20, 2013.



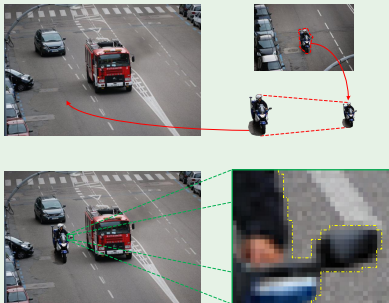
Universidade de Vigo

Problem

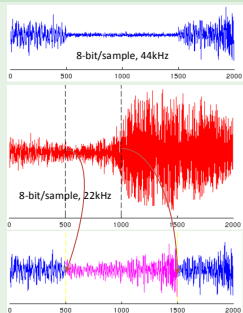
The presence of forged digital contents in the media, social networks or flowing through the Internet has become prevalent these days.

Image/audio splicing requires resampling operations:

Uncompressed images

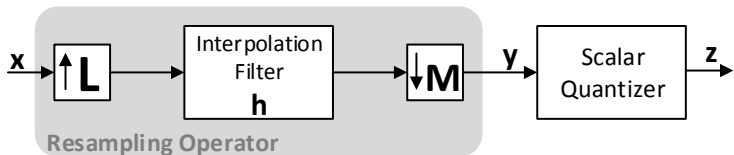


Uncompressed audio signals



BUT, the resampling factor should be constant along the content.

Different Methods for Exposing Resampling Inconsistencies



Analysis in the frequency domain

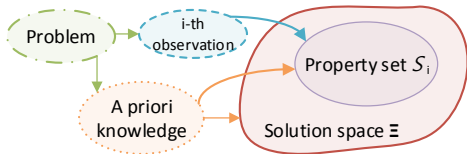
- **Popescu & Farid (2005)**: frequency analysis of probability maps.
- **Kirchner (2008)**: frequency analysis of linear predictor residues.
- **Mahdian & Saic (2008)**: frequency analysis of covariance.

Analysis of the applied quantization

- **Vázquez-Padín & Comesaña (2012)**: approximation to the likelihood function of the quantized resampled signal.

Set-Membership Theory

Set-membership theory is governed by the concept of feasibility:



- $S_i = \{a \in \Xi : a \text{ satisfies } \Psi_i\}$
 Ψ_i : i -th problem constraint

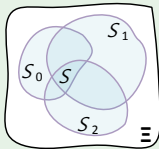
- **Feasible set of solutions:**

$$\mathcal{S} = \cap_{i=0}^{N-1} S_i$$

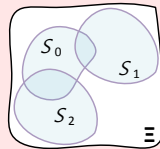
Final goal: find a solution that satisfies all the constraints.

- Given N observations: find $a \in \mathcal{S}$, where $\mathcal{S} = \cap_{i=0}^{N-1} S_i$.

$a \in \mathcal{S}$ (a feasible solution exists)

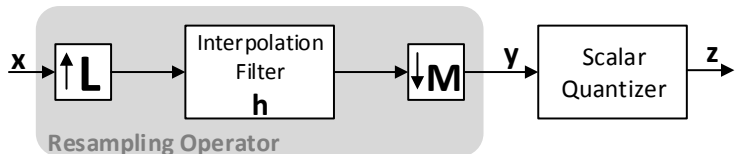


$\mathcal{S} = \emptyset$ (no feasible solution exists)



Set-Membership Formulation of the Resampling Problem

Block diagram representation of the resampling process by $\xi \triangleq \frac{L}{M}$:



A priori knowledge:

- $\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^{N_x} : x_{\min} \leq x_i \leq x_{\max}, i = 0, \dots, N_x - 1\}$
- $\mathcal{H} = \{\mathbf{h} \in \mathbb{R}^{N_h} : h_{\min} \leq h_i \leq h_{\max}, i = 0, \dots, N_h - 1\}$
- Solution space: $\Xi = \mathcal{X} \times \mathcal{H}$
- From the quantization: $z_i - \frac{\Delta}{2} < y_i \leq z_i + \frac{\Delta}{2}$, for $i = 0, \dots, N_z - 1$

From each i -th observation:

$$\mathcal{S}_i = \left\{ (\mathbf{x}, \mathbf{h}) \in \Xi : -\frac{\Delta}{2} < \mathbf{x}_i^T \mathbf{h} - z_i \leq \frac{\Delta}{2} \right\}$$

Feasibility Problem

Given:

- a vector of observations $\mathbf{z}$ of length N_z ,
- a candidate resampling factor $\xi_c \triangleq \frac{L_c}{M_c}$,
- the length of the interpolation filter N_h .

$$\begin{aligned} &\text{find} && \mathbf{x}, \mathbf{h}, \\ &\text{subject to} && \mathbf{x} \in \mathbb{R}^{N_x}, \mathbf{h} \in \mathbb{R}^{N_h}, \\ &&& x_{\min} \leq x_i \leq x_{\max}, && i = 0, \dots, N_x - 1, \\ &&& h_{\min} \leq h_j \leq h_{\max}, && j = 0, \dots, N_h - 1, \\ &&& -\frac{\Delta}{2} < \mathbf{x}_k^T \mathbf{h} - z_k \leq \frac{\Delta}{2}, && k = 0, \dots, N_z - 1. \end{aligned}$$

- **No feasible solution is found:** $\mathbf{z}$ has not been resampled by ξ_c .
- **A feasible solution is found:** $\mathbf{z}$ might have been resampled by ξ_c .

Practical Algorithms

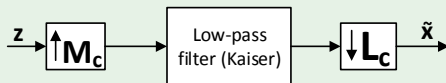
Optimal Value Ellipsoid (OVE) algorithm

Cannot be applied! The feasibility problem has bilinear constraints.

Nonlinear programming algorithms

- **Global optimization techniques:** branch-and-bound strategy.
 - Difficulties handling large-scale problems.
- **Local optimization techniques:** interior-point method.
 - A starting point is needed for $\mathbf{x}$ and $\mathbf{h}$.

Derivation of the starting point: $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{h}}$.



$$\mathbf{z} = Q_{\Delta}(\mathbf{X}\mathbf{h}) \approx \mathbf{X}\mathbf{h}$$

From $\tilde{\mathbf{x}}$, we derive $\tilde{\mathbf{X}}$:

$$\tilde{\mathbf{h}} = \tilde{\mathbf{X}}^+ \mathbf{z}$$

Experimental Validation

1. Synthetic signals: x

- Discrete uniform distribution.
- 8-bit precision samples.
- $x_{\min} = 0, x_{\max} = 2^8 - 1$.

2. Audio signals: x

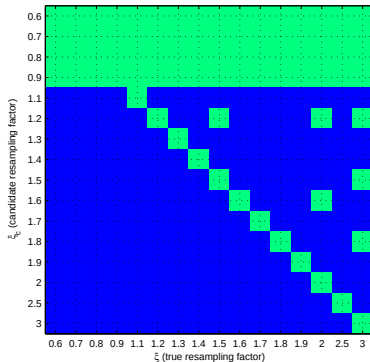
- 100 uncompressed audio files.
- 16-bit precision samples.
- $x_{\min} = 0, x_{\max} = 2^{16} - 1$

Interpolation filters: h

	Scenario 1	Scenario 2
$\xi < 1$	Kaiser, $N_h = 2M + 1$	Kaiser, $N_h = 4M + 1$
$\xi > 1$	Linear, $N_h = 2L + 1$	Cubic, $N_h = 4L + 1$

Quantization: $\Delta = 1$

Performance Analysis with Synthetic Signals



- 1 For $\xi_c < 1$, there is always a feasible solution.
There is not sufficient information (prior or observed).
- 2 For $\xi_c = \xi$, a feasible solution is always found.
Reliable solutions are obtained.
- 3 For $\xi_c \leq \xi$ ($\xi > 1$), multiple candidates are feasible.

The set of possible candidates

$$\mathcal{C} = \left\{ \frac{L_c}{M_c} : \frac{L_c}{M_c} \leq \xi, (L_c = kL) \wedge (M_c > kM), k \in \mathbb{N}^+ \right\}$$

where $L_c \in \mathbb{N}^+$ and $M_c \in \mathbb{N}^+$ are coprime.

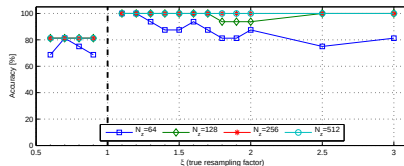
Accuracy analysis for different number of observations

Measuring accuracy

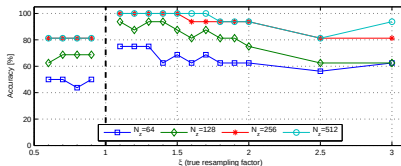
$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}$$

- TP: a feasible solution is found and $\xi_c \in \mathcal{C}$.
- TN: no feasible solution is achieved and $\xi_c \notin \mathcal{C}$.

Scenario 1



Scenario 2



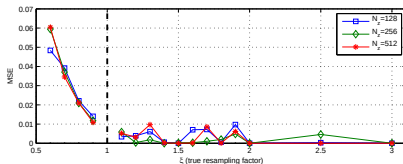
MSE analysis for different number of observations

After checking feasibility, a pair of estimates $\hat{\mathbf{x}}$ and $\hat{\mathbf{h}}$ can be computed.

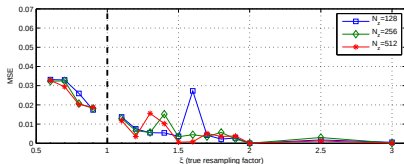
Optimization problem

$$\begin{aligned} & \text{minimize} && \|\mathbf{X}\mathbf{h} - \mathbf{z}\|_2^2, \\ & \text{subject to} && \mathbf{x} \in \mathbb{R}^{N_x}, \mathbf{h} \in \mathbb{R}^{N_h}, \\ & && x_{\min} \leq x_i \leq x_{\max}, && i = 0, \dots, N_x - 1, \\ & && h_{\min} \leq h_j \leq h_{\max}, && j = 0, \dots, N_h - 1, \\ & && -\frac{\Delta}{2} < \mathbf{x}_k^T \mathbf{h} - z_k \leq \frac{\Delta}{2}, && k = 0, \dots, N_z - 1, \end{aligned}$$

MSE of $\hat{\mathbf{h}}$ in **Scenario 1**



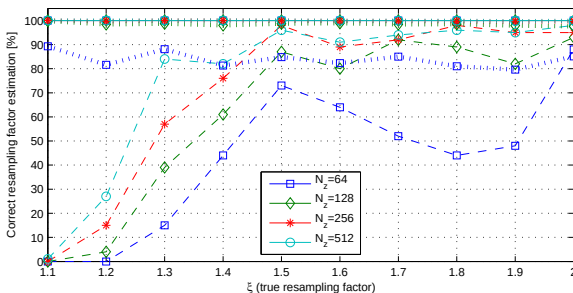
MSE of $\hat{\mathbf{h}}$ in **Scenario 2**



Performance analysis with real audio signals

Comparative results

- **Popescu & Farid**: frequency-based analysis (dashed lines).
- **Vázquez-Padín & Comesaña**: quantization analysis (solid lines).
- **This work**: set-membership approach (dotted lines).



Conclusions and Future Work

Conclusions

- The set-membership identification of resampled signals provides solutions that do not violate any constraint of the problem.
- Answers provided by the forensic analyst will be only based on the observed data and the a priori knowledge about the problem.

Future Work

- Derivation of a probabilistic bound on the number of samples needed to discard a candidate resampling factor.
- Analysis of different interpolation strategies, i.e., content-dependent, non-linear, etc.
- Two-dimensional extension of the set-membership approach for testing uncompressed images.

Thanks for your attention.

Questions?

David Vázquez-Padín
dvazquez@gts.uvigo.es