

EXPOSING ORIGINAL AND DUPLICATED REGIONS USING SIFT FEATURES AND RESAMPLING TRACES

David Vázquez-Padín¹, Fernando Pérez-González^{1,2,3}

1. Signal Theory and Communications Department, University of Vigo, Vigo 36310
2. Electrical and Computer Engineering Department, University of New Mexico, Albuquerque
3. GRADIANT, Vigo, Spain.

Atlantic City, New Jersey, USA. October 25, 2011.

Outline

- 1 Introduction
- 2 Analysis of complementary forensic tools
- 3 Model Description
- 4 Experimental Results
- 5 Conclusions

Introduction

Problem

- A lot of powerful and intuitive image editing tools facilitate the manipulation and alteration of digital images.
- A common type of digital image forgery is the duplication of a region in the same image to conceal something in a captured scene.

Example of image editing software



Image Doctor



Original image



Modified image

Introduction (II)

Example of image tampering



Original Image



Tampered Image

Passive image forensics techniques

- Detection of lighting inconsistencies
- Detection of double compression
- Detection of duplicated regions
- Detection of resampling factor inconsistencies

Analysis of complementary forensic tools

Detection of region duplication

- **Fridrich et al. (2003)**: compare quantized DCT coefficients of overlapped windows.
- **Pan & Lyu (2010)**, **Amerini et al. (2010)**: compare SIFT features and estimate the transformation.

Detection of resampling factor inconsistencies

- **Mahdian & Saic (2008)**: apply a differentiator to the image, use a Radon transform and then process the unidimensional covariance.
- **Kirchner (2008)**: uses the variance of prediction residuals to describe periodic artifacts.
- **Vázquez-Padín et al. (2010)**: apply a statistical test for study the presence of cyclostationarity.

SIFT-based methods vs. Resampling-based methods

SIFT-based methods

Advantages

- Can detect copy-move forgeries.
- Detection results are not block-based.

Drawbacks

- Cannot differentiate original and duplicated regions.
- High dependency between detection performance and image content.

Resampling-based methods

Advantages

- Can identify original and duplicated regions.
- Low dependency between detection performance and image content.

Drawbacks

- Cannot detect copy-move forgeries.
- Block-based analysis for detection.

SIFT-based methods vs. Resampling-based methods

SIFT-based methods

Advantages

- Can detect copy-move forgeries.
- Detection results are not block-based.

Drawbacks

- Cannot differentiate original and duplicated regions.
- High dependency between detection performance and image content.

Resampling-based methods

Advantages

- Can identify original and duplicated regions.
- Low dependency between detection performance and image content.

Drawbacks

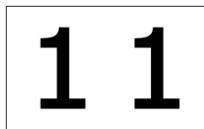
- Cannot detect copy-move forgeries.
- Block-based analysis for detection.

SIFT-based methods vs. Resampling-based methods

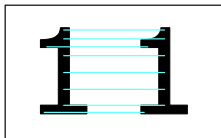
Original image



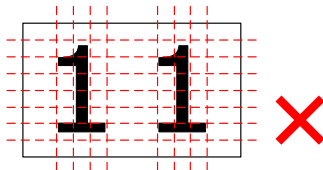
Modified image



SIFT-based methods



Resampling-based methods



Detection of copy-move forgeries.

SIFT-based methods vs. Resampling-based methods

SIFT-based methods

Advantages

- Can detect copy-move forgeries.
- Detection results are not block-based.

Drawbacks

- Cannot differentiate original and duplicated regions.
- High dependency between detection performance and image content.

Resampling-based methods

Advantages

- Can identify original and duplicated regions.
- Low dependency between detection performance and image content.

Drawbacks

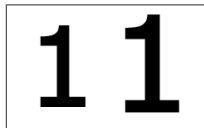
- Cannot detect copy-move forgeries.
- Block-based analysis for detection.

SIFT-based methods vs. Resampling-based methods

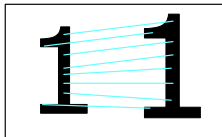
Original image



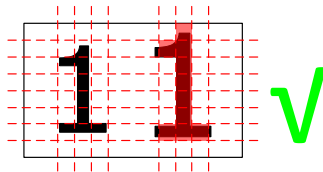
Modified image



SIFT-based methods



Resampling-based methods



Identification of original and duplicated regions.

SIFT-based methods vs. Resampling-based methods

SIFT-based methods

Advantages

- Can detect copy-move forgeries.
- Detection results are not block-based.

Drawbacks

- Cannot differentiate original and duplicated regions.
- High dependency between detection performance and image content.

Resampling-based methods

Advantages

- Can identify original and duplicated regions.
- Low dependency between detection performance and image content.

Drawbacks

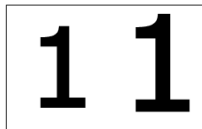
- Cannot detect copy-move forgeries.
- Block-based analysis for detection.

SIFT-based methods vs. Resampling-based methods

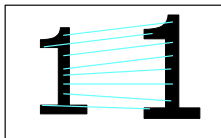
Original image



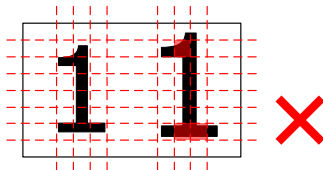
Modified image



SIFT-based methods



Resampling-based methods



Non block-based processing.

SIFT-based methods vs. Resampling-based methods

SIFT-based methods

Advantages

- Can detect copy-move forgeries.
- Detection results are not block-based.

Drawbacks

- Cannot differentiate original and duplicated regions.
- High dependency between detection performance and image content.

Resampling-based methods

Advantages

- Can identify original and duplicated regions.
- Low dependency between detection performance and image content.

Drawbacks

- Cannot detect copy-move forgeries.
- Block-based analysis for detection.

SIFT-based methods vs. Resampling-based methods

Original image



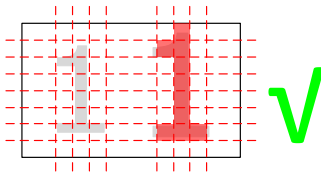
Modified image



SIFT-based methods



Resampling-based methods

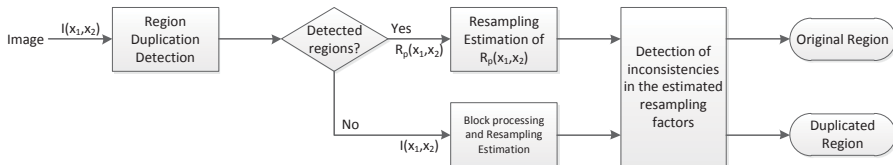


Low dependency between detection performance and image content.

Model description

Notation:

- $I(x_1, x_2)$ represents one channel of the image.
- $R_p(x_1, x_2)$ denotes the p th detected region.



Block Diagram of the proposed image forensics analysis tool.

A SIFT-based method for region duplication

Algorithm

- 1 Obtain SIFT features using Lowe's algorithm:

$$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n : \mathbf{x}_k = (x_1, x_2)\}$$

$$\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}.$$

- 2 Compute Euclidean distance between each pair $(d_i, d_j), i \neq j$:

$$S = \{s_l = \|\mathbf{d}_i - \mathbf{d}_j\|_2 : 1 \leq j \leq n\}.$$

- 3 Compare the distance between the closest neighbor: $\frac{s_1}{s_2} < \Upsilon$.

- 4 Hierarchical clustering.

- 5 Estimate the relation between matched regions:
$$\begin{bmatrix} \mathbf{x}_q^T \\ 1 \end{bmatrix} = \mathbf{H}_{qr} \begin{bmatrix} \mathbf{x}_r^T \\ 1 \end{bmatrix}.$$

A SIFT-based method for region duplication

Algorithm

- 1 Obtain SIFT features using Lowe's algorithm:

$$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n : \mathbf{x}_k = (x_1, x_2)\}$$

$$\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}.$$

- 2 Compute Euclidean distance between each pair $(d_i, d_j), i \neq j$:

$$S = \{s_l = \|\mathbf{d}_i - \mathbf{d}_j\|_2 : 1 \leq j \leq n\}.$$

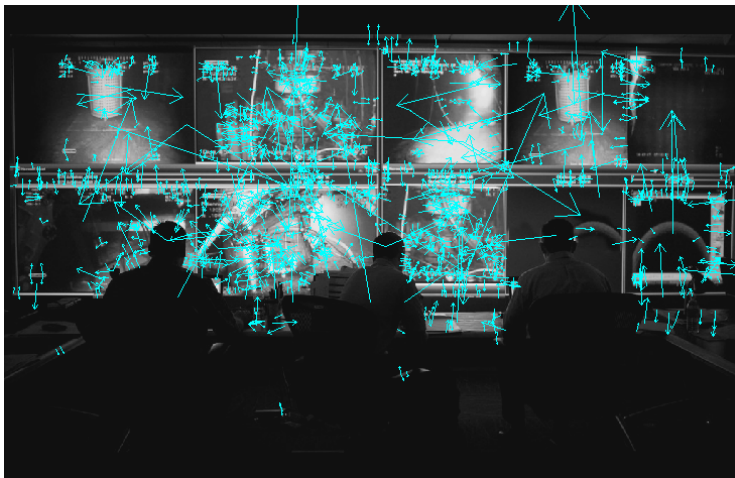
- 3 Compare the distance between the closest neighbor: $\frac{s_1}{s_2} < \Upsilon$.

- 4 Hierarchical clustering.

- 5 Estimate the relation between matched regions:
$$\begin{bmatrix} \mathbf{x}_q^T \\ 1 \end{bmatrix} = \mathbf{H}_{qr} \begin{bmatrix} \mathbf{x}_r^T \\ 1 \end{bmatrix}.$$



A SIFT-based method for region duplication



A SIFT-based method for region duplication

Algorithm

- 1 Obtain SIFT features using Lowe's algorithm:

$$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n : \mathbf{x}_k = (x_1, x_2)\}$$

$$\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}.$$

- 2 Compute Euclidean distance between each pair $(d_i, d_j), i \neq j$:

$$S = \{s_l = \|\mathbf{d}_i - \mathbf{d}_j\|_2 : 1 \leq j \leq n\}.$$

- 3 Compare the distance between the closest neighbor: $\frac{s_1}{s_2} < \Upsilon$.

- 4 Hierarchical clustering.

- 5 Estimate the relation between matched regions:
$$\begin{bmatrix} \mathbf{x}_q^T \\ 1 \end{bmatrix} = \mathbf{H}_{qr} \begin{bmatrix} \mathbf{x}_r^T \\ 1 \end{bmatrix}.$$



A SIFT-based method for region duplication



A SIFT-based method for region duplication

Algorithm

- 1 Obtain SIFT features using Lowe's algorithm:

$$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n : \mathbf{x}_k = (x_1, x_2)\}$$

$$\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}.$$

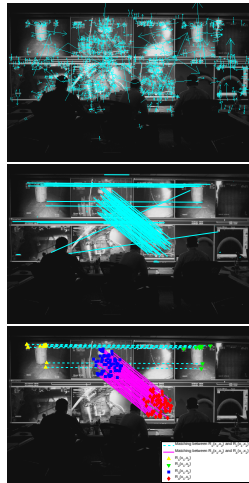
- 2 Compute Euclidean distance between each pair $(d_i, d_j), i \neq j$:

$$S = \{s_l = \|\mathbf{d}_i - \mathbf{d}_j\|_2 : 1 \leq j \leq n\}.$$

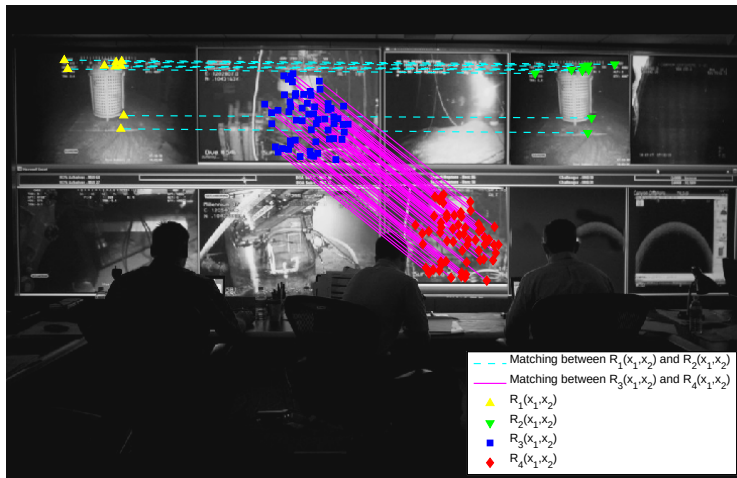
- 3 Compare the distance between the closest neighbor: $\frac{s_1}{s_2} < \Upsilon$.

- 4 Hierarchical clustering.

- 5 Estimate the relation between matched regions:
$$\begin{bmatrix} \mathbf{x}_q^T \\ 1 \end{bmatrix} = \mathbf{H}_{qr} \begin{bmatrix} \mathbf{x}_r^T \\ 1 \end{bmatrix}.$$



A SIFT-based method for region duplication



A SIFT-based method for region duplication

Algorithm

- 1 Obtain SIFT features using Lowe's algorithm:

$$\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n : \mathbf{x}_k = (x_1, x_2)\}$$

$$\mathcal{D} = \{\mathbf{d}_1, \dots, \mathbf{d}_n\}.$$

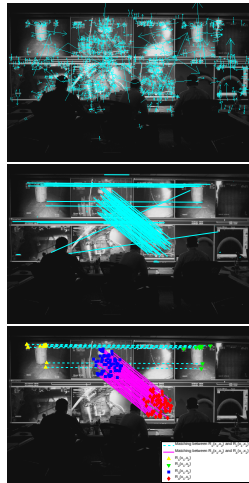
- 2 Compute Euclidean distance between each pair $(d_i, d_j), i \neq j$:

$$S = \{s_l = \|\mathbf{d}_i - \mathbf{d}_j\|_2 : 1 \leq j \leq n\}.$$

- 3 Compare the distance between the closest neighbor: $\frac{s_1}{s_2} < \Upsilon$.

- 4 Hierarchical clustering.

- 5 Estimate the relation between matched regions:
$$\begin{bmatrix} \mathbf{x}_q^T \\ 1 \end{bmatrix} = \mathbf{H}_{qr} \begin{bmatrix} \mathbf{x}_r^T \\ 1 \end{bmatrix}.$$



A resampling-based method to reveal tampered regions

Algorithm

- ❶ Pad with zeros the non-square regions.
- ❷ Apply the statistical test:
 - From the data $R_i(x_1, x_2)$, build up a cyclostationary vector $\hat{\mathbf{c}}_{xx}$, for a set of lags τ .
 - Estimate the covariance matrix $\hat{\mathbf{\Sigma}}_{xx}$.
 - Compute the test statistic
$$\mathcal{T}_{xx} = N^2 \hat{\mathbf{c}}_{xx}^H \hat{\mathbf{\Sigma}}_{xx}^{-1} \hat{\mathbf{c}}_{xx}.$$
 - Set Γ based on the probability of false alarm P_F .
 - Consider a cyclic frequency pair if $\mathcal{T}_{xx} > \Gamma$.
- ❸ Estimate the resampling factor $\hat{\mathbf{N}}_s = \frac{2\pi}{\alpha}$.
- ❹ Analysis of the resampling factor to detect inconsistencies.

A resampling-based method to reveal tampered regions

Algorithm

- ❶ Pad with zeros the non-square regions.
- ❷ Apply the statistical test:
 - From the data $R_i(x_1, x_2)$, build up a cyclostationary vector $\hat{\mathbf{c}}_{xx}$, for a set of lags τ .
 - Estimate the covariance matrix $\hat{\Sigma}_{xx}$.
 - Compute the test statistic
$$\mathcal{T}_{xx} = N^2 \hat{\mathbf{c}}_{xx}^H \hat{\Sigma}_{xx}^{-1} \hat{\mathbf{c}}_{xx}.$$
 - Set Γ based on the probability of false alarm P_F .
 - Consider a cyclic frequency pair if $\mathcal{T}_{xx} > \Gamma$.
- ❸ Estimate the resampling factor $\hat{\mathbf{N}}_s = \frac{2\pi}{\alpha}$.
- ❹ Analysis of the resampling factor to detect inconsistencies.



A resampling-based method to reveal tampered regions



(a) Original region

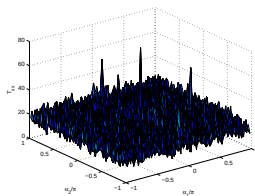


(b) Duplicated region

A resampling-based method to reveal tampered regions

Algorithm

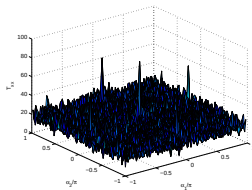
- 1 Pad with zeros the non-square regions.
- 2 Apply the statistical test:
 - From the data $R_i(x_1, x_2)$, build up a cyclostationary vector $\hat{\mathbf{c}}_{xx}$, for a set of lags τ .
 - Estimate the covariance matrix $\hat{\Sigma}_{xx}$.
 - Compute the test statistic
$$\mathcal{T}_{xx} = N^2 \hat{\mathbf{c}}_{xx}^H \hat{\Sigma}_{xx}^{-1} \hat{\mathbf{c}}_{xx}.$$
 - Set Γ based on the probability of false alarm P_F .
 - Consider a cyclic frequency pair if $\mathcal{T}_{xx} > \Gamma$.
- 3 Estimate the resampling factor $\hat{N}_s = \frac{2\pi}{\alpha}$.
- 4 Analysis of the resampling factor to detect inconsistencies.



A resampling-based method to reveal tampered regions



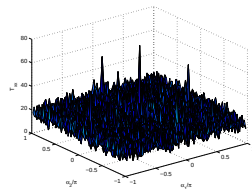
(a) Original region



(c) Test statistic



(b) Duplicated region



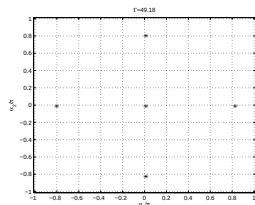
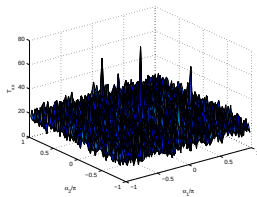
(d) Test statistic

A resampling-based method to reveal tampered regions

Algorithm

- 1 Pad with zeros the non-square regions.
- 2 Apply the statistical test:
 - From the data $R_i(x_1, x_2)$, build up a cyclostationary vector $\hat{\mathbf{c}}_{xx}$, for a set of lags τ .
 - Estimate the covariance matrix $\hat{\Sigma}_{xx}$.
 - Compute the test statistic

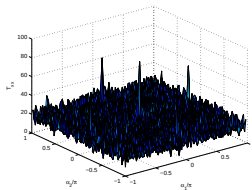
$$\mathcal{T}_{xx} = N^2 \hat{\mathbf{c}}_{xx}^H \hat{\Sigma}_{xx}^{-1} \hat{\mathbf{c}}_{xx}.$$
 - Set Γ based on the probability of false alarm P_F .
 - Consider a cyclic frequency pair if $\mathcal{T}_{xx} > \Gamma$.
- 3 Estimate the resampling factor $\hat{N}_s = \frac{2\pi}{\alpha}$.
- 4 Analysis of the resampling factor to detect inconsistencies.



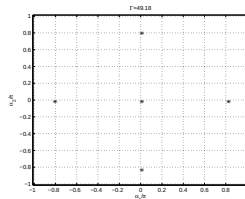
A resampling-based method to reveal tampered regions



(a) Original region



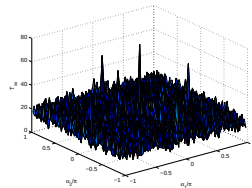
(c) Test statistic



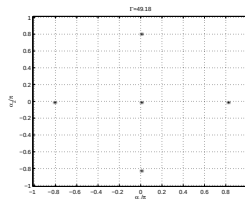
(e) Detected cyclic frequencies



(b) Duplicated region



(d) Test statistic

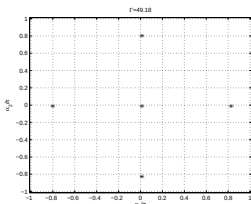
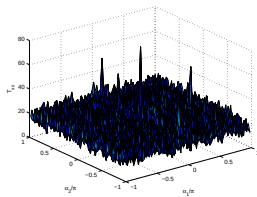


(f) Detected cyclic frequencies

A resampling-based method to reveal tampered regions

Algorithm

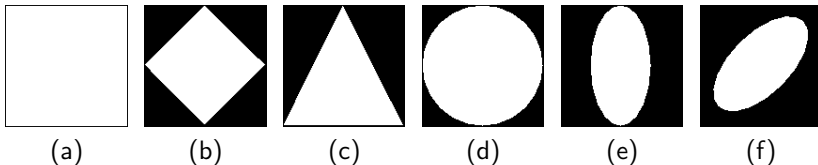
- 1 Pad with zeros the non-square regions.
- 2 Apply the statistical test:
 - From the data $R_i(x_1, x_2)$, build up a cyclostationary vector $\hat{\mathbf{c}}_{xx}$, for a set of lags τ .
 - Estimate the covariance matrix $\hat{\Sigma}_{xx}$.
 - Compute the test statistic $\mathcal{T}_{xx} = N^2 \hat{\mathbf{c}}_{xx}^H \hat{\Sigma}_{xx}^{-1} \hat{\mathbf{c}}_{xx}$.
 - Set Γ based on the probability of false alarm P_F .
 - Consider a cyclic frequency pair if $\mathcal{T}_{xx} > \Gamma$.
- 3 Estimate the resampling factor $\hat{N}_s = \frac{2\pi}{\alpha}$.
- 4 Analysis of the resampling factor to detect inconsistencies.



Experimental Results

Simulations details:

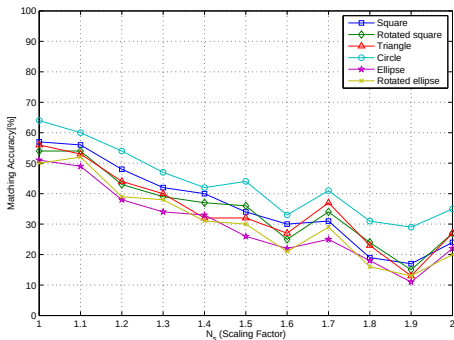
- Image database: 100 non-compressed RGB-images, size 1024×1024 , (only G component).
- Original resampling factor: $\rho_{\text{orig}} = 2$ (due to CFA interpolation).
- Region duplication using 6 different masks.
- Scaling factor for duplicated regions $1 \leq N_s \leq 2$.
- Size of tampered regions: 128×128 pixels.



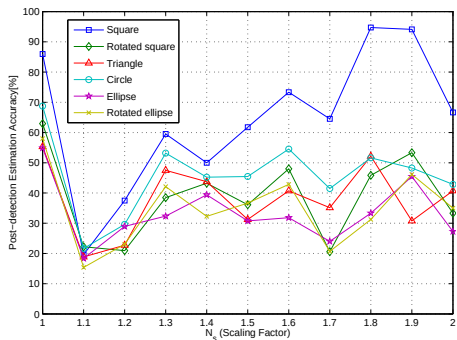
Different masks used to test the performance of the proposed forensic tool.

Experimental Results (II)

Detection results using the SIFT-based method:



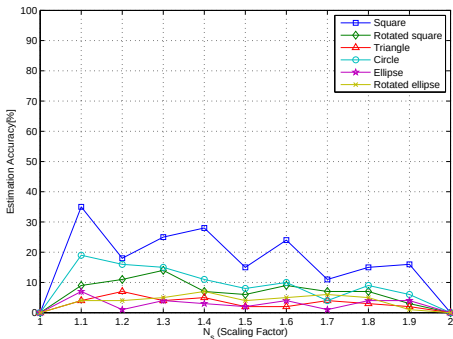
Matching Accuracy



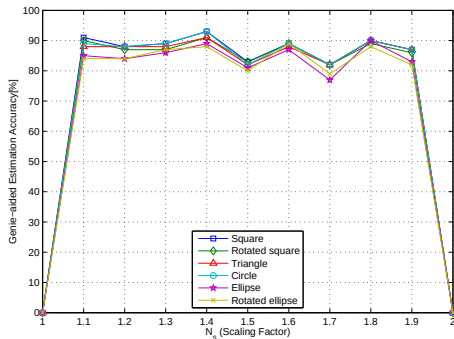
Post-detection Estimation Accuracy

Experimental Results (III)

Detection results using the resampling-based method:



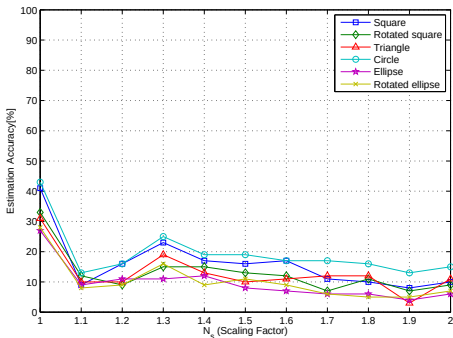
Estimation Accuracy



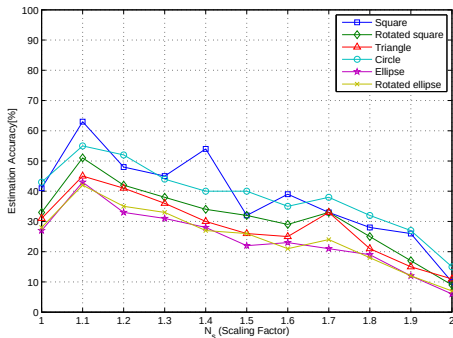
Genie-aided Estimation Accuracy

Experimental Results (IV)

Detection results using both methods:



Estimation Accuracy (SIFT-based method)



Estimation Accuracy

Conclusions

A new scheme for image forensics analysis has been proposed:

- Distinguish manipulated regions of an image from the original ones.
- SIFT-based method provides the location of both regions.
- Resampling-based method allows us to identify the original and the duplicated region.
- Better results are achieved combining both approaches.

Future work:

- Evaluate the same solution using JPEG compressed images.
- Study the estimation of the resampling factor on a non-square area.
- Fuse more image forensics techniques to get better performance and a deeper analysis.

Thanks for your attention.

Questions?

David Vázquez-Padín
dvazquez@gts.uvigo.es