2018 MediFor Challenge

July 27, 2018

Jonathan Fiscus (PI)*
Dr. Haiying Guan*
Andrew Delgado*
Timothee Kheyrkhah*
Dr. Yooyoung Lee*
Daniel Zhou*
Dr. Amy Yates+
August Pereira*

* Multimodal Information Group
+ Image Group
Information Access Division
Information Technology Laboratory
National Institute of Standards and Technology (NIST)
Outline

• MediFor Evaluation Tasks
• MediFor Data Sets
• NIST Evaluation Software: MediScore
• NIST Evaluation Scoring Server
MediFor Evaluation Infrastructure

System Algorithm

NIST Evaluation Dataset (without reference)
Index

Green: System input
Yellow: Performer modules
Blue: NIST Evaluation modules

System Output
Output File (ProbeStatus)
Mask
Filtering Results
Ordered small set
Graph

Scoring Module
Scoring for Detection
Scoring for Localization
Scoring for Provenance Filtering
Scoring for Provenance Graph Building

Performance Report
ROC
Mask Score 0.58
Recall_{first 100}
Graph Similarity Sim_{VEO}(nodes+links)
Cost Function
Cost Value

NIST Evaluation Reference
Output File (ProbeStatus)
Mask
Filtering Results
Ordered small set
Graph

NIST Evaluation Dataset
Index
Media Forensic Challenge (MFC) Evaluation Tasks

• Image Manipulation Detection
• Image Manipulation Localization
• Splice Detection and Localization
• Video Manipulation Detection
• Video Temporal Localization
• Provenance Filtering
• Provenance Graph Building
• Event Verification
Image and Splice Manipulation Detection and Localization

(All images, graphs, and charts in all slides are original works created under contract on the MediFor Program)

**System Input**
Image Detection and Localization
Image(s) + (Metadata)

**System Output**
Confidence score 97.86

**Metrics**
Receiver Operating Characteristic (ROC)
Area Under the Curve (AUC)
Correct Detection (CD) at False Alarm Rate 5%

**Image/Splice Detection and Localization Analytic System**

**Splice Detection and Localization**
Donor image

**System output probe mask**
Manipulated image
Matthews Correlation Coefficient (MCC)

**System output donor mask**
Donor image; MCC

MediFor Program
Video Manipulation Detection and Temporal Localization

• Video Detection metrics
  • Receiver Operating Characteristic (ROC)
  • Area Under the Curve (AUC)
  • Correct Detection (CD) at False Alarm Rate (FAR) of 5%

• Video Temporal Localization – Matthew Correlation Coefficient (MCC)

Figure: Video Temporal Detection and Localization
Provenance Filtering and Graph Building

**System Input**

- **Probe Image**

**System Output**

- **Filtering:** A set of N images with confidence score

- **Graph Building:** A provenance graph

**Algorithm**

**World Image Set** (a large collection of images from internet, ≈1M)

**Oracle Set** (a small collection with all related nodes of the given probe, < 300 for each probe)

**Metrics**

**Filtering:**

Recall

\[ \text{recall} = \frac{|\{\text{relevant}\} \cap \{\text{retrieved}\}|}{|\{\text{relevant}\}|} \]

**Graph Building:**

Generalized F-measure:

\[
\text{sim}_{\text{NLO}}(G_r, G_s) = 2 \frac{|V_r \cap V_s| + |E_r \cap E_s|}{|V_r| + |V_s| + |E_r| + |E_s|}
\]

...
Event Verification (1)

• Task Definition: Given a collection of images and videos from the event, determine if a probe is from the claimed event.

• MFC18 Events
  • 12 events: 6 hurricane, 3 air show, 3 others

Event Verification (2)

• System Input
  • Image and event name pair

• Metrics
  • Receiver Operating Characteristic (ROC), Area Under the Curve (AUC)
  • Correct Detection (CD) at False Alarm Rate (FAR) of 5%

• Training Data
  • N (N ≈ 200) images for each events. 12 events in MFC18: 6 hurricane, 3 air show, 3 others
  • Test and Training images’ camera IDs are different.

• Testing Data
  • M (N ≈ 50) images for each event, then cartesian production with all event name to create pairs.
Outline

✓ MediFor Evaluation Tasks

➢ MediFor Data Sets

• NIST Evaluation Software: MediScore
• NIST Evaluation Scoring Server
Media Forensic Data Characteristics

• **Very diverse topics and tasks**
  • Close collaborations among DARPA, TA1, TA3, and NIST
  • Collection requirements: [https://mediforprogram.com/wiki/display/MEDIFOR/Year+2+Evaluation+and+Data+Needs](https://mediforprogram.com/wiki/display/MEDIFOR/Year+2+Evaluation+and+Data+Needs).

• **Active and generic online metadata collection**
  • Post ground-truth annotation is difficult or impossible.
  • The metadata was collected during the manipulation process with Journaling Tool – high cost.
  • Across different manipulation software and approaches.

• **High intrinsic dimensionalities**
  • Dimension space: image/video, metadata (EXIF, camera ID, captions, association among information units), manipulation (operations, filters, algorithms, and their history orders), semantic meaning.
  • Rich metadata are collected, annotated, recorded, selected, and sampled for evaluation analysis.

• **Data release control mechanism**
  • Camera selection
  • Training and testing data selection
  • Evaluation dataset generation
MediFor Data History

Kick-off 2016 Dataset

Auto Journaling Tool (JT)
Nimble Challenge 2017

New specified data (CGI, Recapture, Video, ...)
Extended JT, AutoJT (Complex)
MediaForensic Challenge 2018
Evaluation Dataset Production Infrastructure

Journaling Tool (JT)

Human Journals

Selection Module

Extended Journals

TestMaker

Sampling Module

Auto Journals

Evaluation Datasets

Forensic Probe Imagery

World Imagery

Test Index

Reference + Masks + Metadata + Journals

Human Manipulator

Camera Fingerprint Training Data

World image

High Provenance

Automated Manipulations

AutoJT BatchJT
MFC18 General Datasets Overview (Image and Video)

Probe dataset
- Manipulation/splice/Provenance
  - High Provenance (HP) -Probe
  - Human Manipulations
  - Extended Manipulations
  - Auto Manipulations

World dataset
- Splice
- Provenance
- PAR Journal images (Base, Donor, Intermediate)

Resource/Training dataset
- Camera Fingerprint Training Data
- Dev. Data sets
- 1/3 of test data ground-truth

Reference dataset

Human Manipulations with High Provenance (HP) -Probe

Extended Manipulations

Auto Manipulations

PAR Journal images (Base, Donor, Intermediate)

Splice

Provenance

World
NIST Image Data Set Summary

- Each dataset contains the HP (high provenance) image, world image, manipulation journals etc.

<table>
<thead>
<tr>
<th>NIST Data Sets</th>
<th>Image Probe (K)</th>
<th>Image Journal</th>
<th>Release Date</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kick Off</td>
<td>1.1</td>
<td>400</td>
<td>07/2016</td>
<td></td>
</tr>
<tr>
<td>NC17 Dev</td>
<td>3.5</td>
<td>≈ 400</td>
<td>04/2017</td>
<td>NC17 Auto</td>
</tr>
<tr>
<td>NC17 Eval</td>
<td>10</td>
<td>-</td>
<td>04/2017</td>
<td>NC17 Auto</td>
</tr>
<tr>
<td>NC17 Eval Part 1</td>
<td>4</td>
<td>406</td>
<td>06/2017</td>
<td>NC17 Auto</td>
</tr>
<tr>
<td>MFC18 Dev1</td>
<td>5.6</td>
<td>178</td>
<td>12/2017</td>
<td></td>
</tr>
<tr>
<td>MFC18 Dev2</td>
<td>38</td>
<td>432</td>
<td>01/2018</td>
<td>Added MFC18 Auto</td>
</tr>
<tr>
<td>MFC18 Eval</td>
<td>50</td>
<td>-</td>
<td>03/2018</td>
<td>Down Select on Probes</td>
</tr>
<tr>
<td>MFC18 EvalPart1</td>
<td>17</td>
<td>758</td>
<td>03/2018</td>
<td>Down Select on Probes</td>
</tr>
</tbody>
</table>
## NIST Video Data Set Summary

<table>
<thead>
<tr>
<th>NIST Data Sets</th>
<th>Video Journal</th>
<th>Video Probe</th>
<th>Release Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>NC17 Dev</td>
<td>≈23</td>
<td>214</td>
<td>04/2017</td>
</tr>
<tr>
<td>NC17 Eval</td>
<td>-</td>
<td>≈ 1K</td>
<td>04/2017</td>
</tr>
<tr>
<td>NC17 Eval Part 1</td>
<td>47</td>
<td>360</td>
<td>06/2017</td>
</tr>
<tr>
<td>MFC18 Dev1</td>
<td>8</td>
<td>118</td>
<td>12/2017</td>
</tr>
<tr>
<td>MFC18 Dev2</td>
<td>36</td>
<td>27</td>
<td>01/2018</td>
</tr>
<tr>
<td>MFC18 Eval</td>
<td>-</td>
<td>3K</td>
<td>03/2018</td>
</tr>
<tr>
<td>MFC18 EvalPart1</td>
<td>-</td>
<td>1K</td>
<td>03/2018</td>
</tr>
</tbody>
</table>
# NIST Provenance Data Set Summary

<table>
<thead>
<tr>
<th>NIST Data Sets</th>
<th>Image Probe</th>
<th>Image Journal</th>
<th>World</th>
<th>Released Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kick Off</td>
<td>-</td>
<td>400</td>
<td>880</td>
<td>07/2016</td>
</tr>
<tr>
<td>NC17 Dev</td>
<td>3.5K</td>
<td>≈ 400</td>
<td>≈ 115K</td>
<td>04/2017</td>
</tr>
<tr>
<td>NC17 Eval</td>
<td>3K</td>
<td>-</td>
<td>1M</td>
<td>04/2017</td>
</tr>
<tr>
<td>NC17 Eval Part 1</td>
<td>1K</td>
<td>406</td>
<td>(1M)</td>
<td>06/2017</td>
</tr>
<tr>
<td>MFC18 Dev1</td>
<td>1.1K</td>
<td>178</td>
<td>10K</td>
<td>12/2017</td>
</tr>
<tr>
<td>MFC18 Dev2</td>
<td>1.2K</td>
<td>432</td>
<td>10K</td>
<td>01/2018</td>
</tr>
<tr>
<td>MFC18 Eval</td>
<td>27K</td>
<td>-</td>
<td>5M</td>
<td>03/2018</td>
</tr>
<tr>
<td>MFC18 EvalPart1</td>
<td>10K</td>
<td>-</td>
<td>1M</td>
<td>03/2018</td>
</tr>
</tbody>
</table>
Outline

✓ MediFor Evaluation Tasks
✓ MediFor Data Sets
➢ NIST Evaluation Software: MediScore
  • NIST Evaluation Scoring Server
Detection Scorer

• Manipulation (image/video), Splice, Event Verification, Camera tasks

• Metrics
  • Area under (ROC) curve (AUC)
    • T# Number of Targets
    • NT# Number of Non-targets
  • (New) CD (Correct detection)
    • @ False Alarm Rate (FAR) = 0.05
    • The TPR (True Positive Rate) value at FAR = 0.05

Performance

\[
\text{Correct Detection Rate} \% = 100 - \text{False Alarm Rate} \%
\]

\[
\text{AUC} = 0.86 \\
\text{(T#: 600, NT#: 595)}
\]
Image Localization (Mask) Scorer

• Applied to Manipulation, Splice, and Camera fingerprint tasks

• Metrics
  • Matthews Correlation Coefficient (MCC)

\[
MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \in [-1,1]
\]
  • 1 denotes perfect accuracy
  • 0 denotes no correlation between reference and system masks
  • -1 denotes perfect inaccuracy.

• Only evaluates on true targets
Localization Scorer – Selective Scoring

- JPEG2000 allows the localization scorer to evaluate for manipulations at any layer of the pixel
## Selective Scoring – Table of Operations (MFC18)

<table>
<thead>
<tr>
<th>Selective Scoring Name</th>
<th>ForImage</th>
<th>ForVideo</th>
<th>ForDetection</th>
<th>ForLocalization</th>
<th>ForProvFilt</th>
<th>ForProvGB</th>
<th>FilterCommand</th>
</tr>
</thead>
<tbody>
<tr>
<td>PasteSplice</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>-qm &quot;Operation==['PasteSplice']&quot;</td>
</tr>
<tr>
<td>ContentAwareRemove</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation==['ContentAwareFill']&quot;</td>
</tr>
<tr>
<td>Clone</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation==['PasteSampled'] and Purpose==['clone']&quot;</td>
</tr>
<tr>
<td>FaceManipulation</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;OperationArgument==['face']&quot;</td>
</tr>
<tr>
<td>Paste</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation == ['PasteSampled','PasteSplice']&quot;</td>
</tr>
<tr>
<td>PasteSampled</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation==['PasteSampled']&quot;</td>
</tr>
<tr>
<td>Remove</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation==['PasteSampled','ContentAwareFill'] and Purpose==['remove']&quot;</td>
</tr>
<tr>
<td>Crop</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation==['TransformCrop']&quot;</td>
</tr>
<tr>
<td>LocalSharpening</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation==['Sharpening'] and Color!=&quot;&quot;</td>
</tr>
<tr>
<td>GlobalIntensityNormalization</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>N</td>
<td>-qm &quot;Operation==['Normalization']&quot;</td>
</tr>
</tbody>
</table>
(NEW) Video Temporal Localization Task

• Evaluation metrics: MCC

\[ MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}} \in [-1,1] \]

- 1 denotes perfect accuracy, 0 denotes no correlation between reference and system masks, -1 denotes perfect inaccuracy.

• System validation check
  - Valid Video Segment and Audio Segments lists: `[ ]` or `[[x_1, x_2], [x_3, x_4], ...]`
  - New --truncate option, to address inconsistent frameCounts issue.

• Score Output

  Sortable HTML dynamic table

  Interactive Timeline Visualisation
Provenance Graph Filtering Task

• Evaluation metric definition

\[
\text{recall} = \frac{|\{\text{relevant} \} \cap \{\text{retrieved} \}|}{|\{\text{relevant} \}|}
\]

• Metric description
  - The recall of the first N images from the world dataset (≈1M) sorted by ‘confidence score’
  - Only true manipulated probes whose contributors are in the world data set are evaluated.
  - Different value for the depth of retrieval: recall@100, 200, 300.

• Node filtering option
  - Compute each recall for each Node Type.
Provenance Graph Building Task

- **Evaluation metrics:** Graph Similarity and Generalized F-measure
  - **Nodes overlap:** $\text{sim}_{\text{NO}}(G_T, G_S) = 2 \frac{|V_T \cap V_S|}{|V_T| + |V_S|}$
  - **Links overlap:** $\text{sim}_{\text{LO}}(G_T, G_S) = 2 \frac{|E_T \cap E_S|}{|E_T| + |E_S|}$
  - **Nodes and links overlap:** $\text{sim}_{\text{NLO}}(G_T, G_S) = 2 \frac{|V_T \cap V_S| + |E_T \cap E_S|}{|V_T| + |V_S| + |E_T| + |E_S|}$

- **Two graph scoring options:**
  - **Direct Path Limited:** All direct ancestor and descendants of a given probe
  - **Full Graph:** Recursively include all paths connected to the probe and all ancestors and descendants

<table>
<thead>
<tr>
<th>Image border</th>
<th>Correctly included</th>
<th>False alarm image</th>
<th>Omitted provenance image (missed detection)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Link</td>
<td>Correctly linked images</td>
<td>False alarm link</td>
<td>Omitted link</td>
</tr>
</tbody>
</table>

Graph legend

[Image of graph legend: Green, Red, Gray]
Outline

✓ MediFor Evaluation Tasks
✓ NIST Data Sets
✓ NIST Evaluation Software: MediScore
➢ NIST Evaluation Scoring Server
Scoring Server

Updates

- Two new evaluation tasks
  - Camera Fingerprint ID
  - Event Verification
- Backend database for analysis
- System output
- Scoring results
- Storage Management
- Balance hard drive usage (~20 TB so far)

Future

- Analysis Pipeline
- Scoring server refresh

Selective Scorings

- Updates to selective scorings for MFC18 Data
- GlobalIntensityNormalization
- LocalSharpening
- ContentAwareRemove
- LocalAdditionalEffectBlur
- CopyPaste – Video Detection
- Localization only scores on localizable selective scorings

System output

- Scoring results

Storage Management

- Balance hard drive usage (~20 TB so far)
Disclaimer

• All images, graphs, and charts are original works created under contract on the DARPA MediFor Program.

• Certain commercial equipment, instruments, software, or materials are identified in this article in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by NIST, nor is it intended to imply that the equipment, instruments, software or materials are necessarily the best available for the purpose.

• The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.
Thank You for Your Attention!

- NIST Medifor Team: mfc_poc@nist.gov

- MediScore Github: https://github.com/usnistgov/MediScore

- NIST Medifor Evaluation (2017 and 2018):
  https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2018