Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

INTERNATIONAL DOCTORATE IN COMPUTER SCIENCE

PhD candidate: Filippo LM Milotta

Supervisor: Prof. S. Battiato

DEPT. OF MATHEMATICS AND COMPUTER SCIENCE
UNIVERSITY OF CATANIA

CATANIA
MARCH 16, 2018
IMAGE PROCESSING LABORATORY

- Cultural Heritage:
  - 3D Modeling – 3D Scanning – Artificial Mosaics – Video Artifacts – Virtual Tour – Saliency Estimation

- Computer Vision:

- Assistive Technology and Medical Imaging:
  - Food Understanding – Finger Tracking – Braille Converter – Obstacle Avoidance
JOINT OPEN LAB - TIM

WAVE
WIRELESS APPLICATIONS IN MULTI-DEVICE ECOSYSTEMS

Filippo L.M. Milotta - PhD Candidate

16-MAR-2018
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

This dissertation collects all the research work done by the PhD candidate in the Joint Open Lab for Wireless Applications in multi-device Ecosystems (JOL WAVE CATANIA) of TIM Telecom Italia.

In this lab, novel mobile applications based on versatile software are designed and implemented. The development process is performed employing **highly connectable hardware** and devices like smartphones, tablets, cameras, wearable devices, sensors, actuators, smart objects, and interactive screens.
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Three main categories of media are treated in this dissertation: images, videos, and 3D data. More in detail, for images and videos we realized two frameworks: The Social Picture and RECfusion, respectively.

A: ARCA
Automatic Recognition of Color for Archaeology

B: 3D Web Viewers
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

- A: ARCA
  Automatic Recognition of Color for Archaeology
- 2: Image Matching and Retrieval
- 3: The Social Picture
- 4: Social Saliency
- 5: Video Summarization
- 6: 3D Data for Cultural Heritage
- 7: 3D Data for Medical Research
- B: 3D Web Viewers
PART 1 – IMAGES
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

2. Image Matching and Retrieval
3. The Social Picture
4. Social Saliency
5. Video Summarization
6. 3D Data for Cultural Heritage
7. 3D Data for Medical Research
A. ARCA - Automatic Recognition of Color for Archaeology
B. 3D Web Viewers
Datasets of images are usually browsed by means of queries based on tags, hash-tags or geodata (where present).

Queries can be even done using other images and looking for the most similar ones in the dataset. This procedure is called “Content Based Image Retrieval” (CBIR) and is based on Image Matching algorithms.

Content Based Image Retrieval (CBIR) has a consolidated workflow:

1. Keypoints Detection
2. Features extraction from detected keypoints (Descriptor definition)
3. Features matching
4. Geometric verification
5. Inliers extraction

Image Matching and Retrieval

Improvements to CBIR (1/3)

▪ **Back-Projection Verification**
  ▪ Sometimes, geometric verification step of CBIR may fail to guarantee a good image transformation. Then, noise is introduced.
  ▪ Query image is “warped” (deformed) applying the homography estimated by CBIR. Then, we compute Image matching between warped query image and the original query image. Warped query image is back-projected in the original query image. Using the same tolerance on inliers number chosen we check if back-projection is verified.

Image Matching and Retrieval

Improvements to CBIR (2/3)

- **Query Expansion (Re-Query)**
  
  A single step of Image Matching sometimes is not enough. It is possible to match two query images between themselves and then find the nested relationship for match the query image discarded in the first step of Image Matching.

Heatmap Computation

When a reference image is chosen, then a heatmap of matched query images can be computed. The heatmap is updated with relation to query images processed with CBIR workflow and CDVS.

Heatmap is a valuable tool for data analysis and summarization that has been implemented within our framework The Social Picture (TSP).

Image Matching and Retrieval

Visual comparison between 4 CBIR methods

Reference Image

CDVS_BP-2

SURF-8

MSER-2

VLFeat-1
## Image Matching and Retrieval

### Visual comparison between 4 CBIR methods

Table 2.7: Heatmap computation performances comparison of tested methods, sorted by True Positive Rate (TPR), computational time and quality (as defined in Section 2.4).

<table>
<thead>
<tr>
<th>Sorted by</th>
<th>TPR (Decreasing order)</th>
<th>TestID</th>
<th>TPR</th>
<th>Time</th>
<th>Quality</th>
<th>Sorted by</th>
<th>Time</th>
<th>Quality</th>
<th>Sorted by</th>
<th>TPR (Decreasing order)</th>
<th>TestID</th>
<th>TPR</th>
<th>Time</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>TestID</td>
<td></td>
<td>CDVS-2</td>
<td>82.19%</td>
<td>~11 h</td>
<td>★★★★★</td>
<td>CDVS-9</td>
<td>3.94%</td>
<td>~15</td>
<td>★★★★★</td>
<td>CDVS-13</td>
<td>47.73%</td>
<td>~18 h</td>
<td>★★★★★</td>
<td></td>
</tr>
<tr>
<td>CDVS-3</td>
<td>63.11%</td>
<td>~3.5 h</td>
<td>★★★★★</td>
<td>CDVS-5</td>
<td>3.94%</td>
<td>~20</td>
<td>★★★★★</td>
<td>SURF-8</td>
<td>38.28%</td>
<td>~11 d</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS_BP-2</td>
<td>47.73%</td>
<td>~18 h</td>
<td>★★★★★</td>
<td>CDVS-7</td>
<td>3.51%</td>
<td>~20</td>
<td>★★★★★</td>
<td>SURF-7</td>
<td>31.19%</td>
<td>~3 d</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-11</td>
<td>45.12%</td>
<td>~5 h</td>
<td>★★★★★</td>
<td>CDVS-18</td>
<td>1.70%</td>
<td>~20</td>
<td>★★★★★</td>
<td>SURF-6</td>
<td>28.36%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-8</td>
<td>38.28%</td>
<td>~11 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>2.54%</td>
<td>~20</td>
<td>★★★★★</td>
<td>CDVS-13</td>
<td>19.26%</td>
<td>~3 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-7</td>
<td>31.19%</td>
<td>~3 d</td>
<td>★★★★★</td>
<td>CDVS-8</td>
<td>3.51%</td>
<td>~25</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>2.54%</td>
<td>~20</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-2</td>
<td>29.50%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS-10</td>
<td>2.79%</td>
<td>~25</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>2.54%</td>
<td>~20</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS_BP-4</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td>CDVS-9</td>
<td>3.51%</td>
<td>~30</td>
<td>★★★★★</td>
<td>CDVS_BP-4</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-6</td>
<td>26.35%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>CDVS-5</td>
<td>23.62%</td>
<td>~3 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-4</td>
<td>25.44%</td>
<td>~4.5 h</td>
<td>★★★★★</td>
<td>CDVS-13</td>
<td>19.26%</td>
<td>~3 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-3</td>
<td>23.62%</td>
<td>~3 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>12.06%</td>
<td>~3.5 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-1</td>
<td>19.26%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS-7</td>
<td>63.11%</td>
<td>~3.5 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-13</td>
<td>19.26%</td>
<td>~3 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>12.06%</td>
<td>~3.5 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-3</td>
<td>15.87%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS-4</td>
<td>25.44%</td>
<td>~1.5 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS_BP-1</td>
<td>12.05%</td>
<td>~3.5 h</td>
<td>★★★★★</td>
<td>CDVS-11</td>
<td>45.12%</td>
<td>~5 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSER-1</td>
<td>11.57%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>CDVS-2</td>
<td>82.19%</td>
<td>~11 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-4</td>
<td>11.51%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>47.73%</td>
<td>~18 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VLPeat-1</td>
<td>11.21%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>SURF-2</td>
<td>29.50%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSER-4</td>
<td>10.84%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>SURF-7</td>
<td>19.26%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSER-6</td>
<td>10.72%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>SURF-3</td>
<td>15.87%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SURF-5</td>
<td>10.66%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>VLPeat-1</td>
<td>11.21%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSER-5</td>
<td>10.39%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>VLPeat-1</td>
<td>11.21%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MSER-3</td>
<td>9.63%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>VLPeat-1</td>
<td>11.21%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-5</td>
<td>3.94%</td>
<td>~15</td>
<td>★★★★★</td>
<td>SURF-6</td>
<td>26.35%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-6</td>
<td>3.94%</td>
<td>~15</td>
<td>★★★★★</td>
<td>SURF-4</td>
<td>11.51%</td>
<td>~1 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-7</td>
<td>3.51%</td>
<td>~20</td>
<td>★★★★★</td>
<td>CDVS-2</td>
<td>82.19%</td>
<td>~11 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-8</td>
<td>3.51%</td>
<td>~20</td>
<td>★★★★★</td>
<td>CDVS-11</td>
<td>45.12%</td>
<td>~5 h</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-9</td>
<td>3.51%</td>
<td>~30</td>
<td>★★★★★</td>
<td>MSER-5</td>
<td>10.30%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-10</td>
<td>2.79%</td>
<td>~25</td>
<td>★★★★★</td>
<td>MSER-9</td>
<td>9.63%</td>
<td>~2 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS_BP-3</td>
<td>2.54%</td>
<td>~30</td>
<td>★★★★★</td>
<td>SURF-7</td>
<td>31.19%</td>
<td>~3 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDVS-12</td>
<td>1.70%</td>
<td>~20</td>
<td>★★★★★</td>
<td>SURF-6</td>
<td>38.28%</td>
<td>~11 d</td>
<td>★★★★★</td>
<td>CDVS_BP-3</td>
<td>28.47%</td>
<td>~4 h</td>
<td>★★★★★</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

2. Image Matching and Retrieval
3. The Social Picture
4. Social Saliency
5. Video Summarization
6. 3D Data for Cultural Heritage
7. 3D Data for Medical Research

A. ARCA: Automatic Recognition of Color for Archaeology
B. 3D Web Viewers
The Social Picture (TSP) Architecture
TSP Overview of a Collection

The Social Picture

Pisa

Filters

All Day Night All Indoor Outdoor All Large Medium Small

Photo details

Name: 4435130.jpg
Resolution: 3072 x 2048
Coordinates: lat: 43.7224, lon: 10.396
Tag: tower, mausoleum, skyscraper, bell tower, bell, lighthouse, beacon light, pharos, castle

Launch TSNE
VSFM integration and Model definition

**VSFM** is a powerful tool of 3D reconstruction from a set of images, exploiting Structure from Motion (SfM)
- VSFM performs a linear-time incremental SfM method
- VSFM extracts visual features and computes 3D reconstruction
- Output is saved in a N-View Match (NVM) file

[82] Simon et al., “Scene summarization for online image collections” (2007)
Feature Density Maps

Exploiting CPC matrix we can select all the points used as features by a given camera. Through the Model, which is parsed from NVM file, we also know the position of features in the image. Given the positions of visual features, we can define several kinds of maps:

- Density Map (D-map)
- Weighted-Density Map (WD-map)
- Social-Weighted-Density Map (SWD-map)
- Cumulative Map (C-map)
Density Map (D-map)

Visual feature density is obtained quantizing the 2D space of the image and counting how many features are contained in each quantized interval.
Weigthed-Density Map (WD-map)

D-maps can be further refined taking into account a weight for each visual feature.

Weight is equal to the feature-frequency, as computed through the column-wise sum from CPC matrix.
Social-Weighted-Density Map (SWD-map)

We estimate the perspective transformation $T$ between matching images. All the WD-maps of matching images, transformed with their own $T$, are summed to the WD-map of the reference image.

The rationale behind the computation of SWD-maps is that images similar to a reference one might contain information that is not present in the reference image.
3D-points Frequency Representation

The column-wise sum of $CPC$ gives as result how many times each point has been viewed by cameras.

This “frequency of been viewed by a camera” can be used in place of the $color$ vertex in the 3D representation.
Recap
# Feature Density Maps – Examples

<table>
<thead>
<tr>
<th>ID</th>
<th>Image</th>
<th>Features</th>
<th>D-Map</th>
<th>WD-Map</th>
<th>SWD-Map</th>
<th>Blended SWD-Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="features1.png" alt="Features" /></td>
<td><img src="d_map1.png" alt="D-Map" /></td>
<td><img src="wd_map1.png" alt="WD-Map" /></td>
<td><img src="swd_map1.png" alt="SWD-Map" /></td>
<td><img src="blended_swd_map1.png" alt="Blended SWD-Map" /></td>
</tr>
<tr>
<td>16</td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="features16.png" alt="Features" /></td>
<td><img src="d_map16.png" alt="D-Map" /></td>
<td><img src="wd_map16.png" alt="WD-Map" /></td>
<td><img src="swd_map16.png" alt="SWD-Map" /></td>
<td><img src="blended_swd_map16.png" alt="Blended SWD-Map" /></td>
</tr>
<tr>
<td>45</td>
<td><img src="image45.png" alt="Image" /></td>
<td><img src="features45.png" alt="Features" /></td>
<td><img src="d_map45.png" alt="D-Map" /></td>
<td><img src="wd_map45.png" alt="WD-Map" /></td>
<td><img src="swd_map45.png" alt="SWD-Map" /></td>
<td><img src="blended_swd_map45.png" alt="Blended SWD-Map" /></td>
</tr>
<tr>
<td>69</td>
<td><img src="image69.png" alt="Image" /></td>
<td><img src="features69.png" alt="Features" /></td>
<td><img src="d_map69.png" alt="D-Map" /></td>
<td><img src="wd_map69.png" alt="WD-Map" /></td>
<td><img src="swd_map69.png" alt="SWD-Map" /></td>
<td><img src="blended_swd_map69.png" alt="Blended SWD-Map" /></td>
</tr>
<tr>
<td>74</td>
<td><img src="image74.png" alt="Image" /></td>
<td><img src="features74.png" alt="Features" /></td>
<td><img src="d_map74.png" alt="D-Map" /></td>
<td><img src="wd_map74.png" alt="WD-Map" /></td>
<td><img src="swd_map74.png" alt="SWD-Map" /></td>
<td><img src="blended_swd_map74.png" alt="Blended SWD-Map" /></td>
</tr>
<tr>
<td>87</td>
<td><img src="image87.png" alt="Image" /></td>
<td><img src="features87.png" alt="Features" /></td>
<td><img src="d_map87.png" alt="D-Map" /></td>
<td><img src="wd_map87.png" alt="WD-Map" /></td>
<td><img src="swd_map87.png" alt="SWD-Map" /></td>
<td><img src="blended_swd_map87.png" alt="Blended SWD-Map" /></td>
</tr>
</tbody>
</table>
Feature Density Maps – Occlusion
Scene Summarization

We compute a score for each node (camera) in MST

- Score = sum of adjacent edge-weights
- Row-wise sum of CCM

We performed a DFS graph traversal on MST, using the node with the highest score as starting node

[82] Simon et al., “Scene summarization for online image collections” (2007)
TSP – Conclusion

Using VSFM and its 3D reconstruction, we defined new features added to TSP, such as the **3D-points frequency**, and presented two advanced Image Analysis applications: **Scene Summarization (SS) and Feature density-maps**.

![Images of 3D reconstructions and feature density maps](image-url)
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

2. Image Matching and Retrieval
3. The Social Picture
4. Social Saliency
5. Video Summarization
6. 3D Data for Cultural Heritage
7. 3D Data for Medical Research
8. 3D Web Viewers

ARCA - Automatic Recognition of Color for Archaeology
### Social Saliency

**Definition of saliency:**

*spatial regions in the visual field that attract attention*

**Social saliency,** intended as a saliency estimated from a social point-of-view, represents a novel definition of saliency.

Similar concept of saliency are defined as **co-saliency,** **multi-camera saliency** or **likelihood of joint attention.**

<table>
<thead>
<tr>
<th>SWD-Map</th>
<th>Blended SWD-Map</th>
</tr>
</thead>
<tbody>
<tr>
<td>![SWD-Map Image]</td>
<td>![Blended SWD-Map Image]</td>
</tr>
<tr>
<td>![SWD-Map Image]</td>
<td>![Blended SWD-Map Image]</td>
</tr>
<tr>
<td>![SWD-Map Image]</td>
<td>![Blended SWD-Map Image]</td>
</tr>
</tbody>
</table>

---

**References:**


Social Saliency

We tested two methods to learn saliency models: one is based on Support Vector Regression (SVR) and the other one on a Counting Convolutional Neural Network (C-CNN) named Hydra C-CNN. We compared results of linear and non-linear SVR with the ones obtained by Hydra.

Social Saliency – Experimental Results

(a) COR - SVR Linear
(b) AUC - SVR Linear
(c) COR - SVR Gaussian
(d) AUC - SVR Gaussian
(e) COR - Hydra
(f) AUC - Hydra

Social Saliency – Experimental Results
Social Saliency – Experimental Results

ID#1 - Image  
ID#1 - SWD-map  
ID#1 - SVR-Linear  
ID#1 - SVR-Gaussian  
ID#1 - Hydra

ID#21 - Image  
ID#21 - SWD-map  
ID#21 - SVR-Linear  
ID#21 - SVR-Gaussian  
ID#21 - Hydra

ID#50 - Image  
ID#50 - SWD-map  
ID#50 - SVR-Linear  
ID#50 - SVR-Gaussian  
ID#50 - Hydra

ID#53 - Image  
ID#53 - SW0-map  
ID#53 - SVR-Linear  
ID#53 - SVR-Gaussian  
ID#53 - Hydra
Part 1 – Images

Publications


PART 2 – VIDEOS
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

1. Image Matching and Retrieval
2. The Social Picture
3. Social Saliency
4. 3D Web Viewers
5. Video Summarization
6. 3D Data for Cultural Heritage
7. 3D Data for Medical Research

A. ARCA
   Automatic Recognition of Color for Archaeology
Motivations

The automatic processing of video data from many devices, as smartphones, tablets, webcams, surveillance cameras, etc., in the real-time context is not a trivial issue.
Main Aims and Challenges

- Analysis of video streams from multi-source multi-device context
  - Different acquisition formats

- Identification of the scenes of interest through clustering of video sequences
  - Different point-of-views from the many users

- Time tracking of the computed scenes clusters
  - Synchronization is needed
Three kinds of classification

Table 5.2: Validation Results of Popularity Estimation.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Devices</th>
<th>Models</th>
<th>$P_o/P_r$</th>
<th>$P_g/P_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foosball</td>
<td>4</td>
<td>2</td>
<td>1.02</td>
<td>1</td>
</tr>
<tr>
<td>Meeting</td>
<td>2</td>
<td>2</td>
<td>1.01</td>
<td>0.99</td>
</tr>
<tr>
<td>Meeting</td>
<td>4</td>
<td>4</td>
<td>0.99</td>
<td>0.95</td>
</tr>
<tr>
<td>Meeting</td>
<td>5</td>
<td>5</td>
<td>0.89</td>
<td>0.76</td>
</tr>
<tr>
<td>SAgata</td>
<td>7</td>
<td>6</td>
<td>1.05</td>
<td>1</td>
</tr>
<tr>
<td>Magician</td>
<td>6</td>
<td>6</td>
<td>0.73</td>
<td>0.73</td>
</tr>
<tr>
<td>Concert [150]</td>
<td>3</td>
<td>1</td>
<td>1.06</td>
<td>1</td>
</tr>
<tr>
<td>Lecture [150]</td>
<td>3</td>
<td>1</td>
<td>1.05</td>
<td>0.86</td>
</tr>
<tr>
<td>Seminar [150]</td>
<td>3</td>
<td>1</td>
<td>0.62</td>
<td>0.62</td>
</tr>
</tbody>
</table>

RECfusion – Conclusion
Part 2 – Videos

Publications


PART 3 – 3D DATA

Case Studies:
1. Doorway of the Monastery of Benedettini in Catania
2. Morgantina Silver Treasure
3. Kourous of Leontinoi
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

1. Image Matching and Retrieval
2. The Social Picture
3. Social Saliency
4. 3D Web Viewers
5. Video Summarization
6. 3D Data for Cultural Heritage
7. 3D Data for Medical Research

A. ARCA
   Automatic Recognition of Color for Archaeology
Motivation: Cultural Heritage Preservation

Multiple Digital Copies

Digital Restoration

Catalogue and Archive (monitoring purposes, ...)

2014

2015

2016

...
3D Scan of CH – Benedettini’s Door
3D Scan of CH - Kouros
Morgantina silver treasure
Part 3 – 3D Data

Cultural Heritage – Publications


Part 3 – 3D Data

Cultural Heritage – Publications


Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

1. Image Matching and Retrieval
2. The Social Picture
3. Social Saliency
4. 3D Web Viewers
5. Video Summarization
6. 3D Data for Cultural Heritage
7. 3D Data for Medical Research

ARCA: Automatic Recognition of Color for Archaeology
3D Scanning for medical imaging
3D Scanning for medical imaging
3D Scanning for medical imaging

Group 1 and Group 2

First principal component

Second principal component

Group 1 Left
Group 1 Right
Group 2 Left
Group 2 Right

Pre-operation
Post-operation
Part 3 – 3D Data

Medical Research – Publications


APPENDICES
Multi-Device Media Analysis and Summarization for High Bandwidth Connected Environment

Slideshow Structure

1. Image Matching and Retrieval
2. ARCA: Automatic Recognition of Color for Archaeology
3. The Social Picture
4. Social Saliency
5. Video Summarization
6. 3D Data for Cultural Heritage
7. 3D Data for Medical Research
8. 3D Web Viewers
### Image Dataset: ARCA328

328 Images (56,160 samples)

<table>
<thead>
<tr>
<th>Set</th>
<th>Name</th>
<th>RAC</th>
<th>RSC</th>
<th>RCC</th>
<th>SAC</th>
<th>SCC</th>
<th>SCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10R</td>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
<tr>
<td>1</td>
<td>7.5YR</td>
<td><img src="image7.png" alt="Image" /></td>
<td><img src="image8.png" alt="Image" /></td>
<td><img src="image9.png" alt="Image" /></td>
<td><img src="image10.png" alt="Image" /></td>
<td><img src="image11.png" alt="Image" /></td>
<td><img src="image12.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td>10R</td>
<td><img src="image13.png" alt="Image" /></td>
<td><img src="image14.png" alt="Image" /></td>
<td><img src="image15.png" alt="Image" /></td>
<td><img src="image16.png" alt="Image" /></td>
<td><img src="image17.png" alt="Image" /></td>
<td><img src="image18.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td>7.5YR</td>
<td><img src="image19.png" alt="Image" /></td>
<td><img src="image20.png" alt="Image" /></td>
<td><img src="image21.png" alt="Image" /></td>
<td><img src="image22.png" alt="Image" /></td>
<td><img src="image23.png" alt="Image" /></td>
<td><img src="image24.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td>Test Case #1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>Test Case #9</td>
<td><img src="image25.png" alt="Image" /></td>
<td><img src="image26.png" alt="Image" /></td>
<td><img src="image27.png" alt="Image" /></td>
<td><img src="image28.png" alt="Image" /></td>
<td><img src="image29.png" alt="Image" /></td>
<td><img src="image30.png" alt="Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Set</th>
<th>Name</th>
<th>RAM</th>
<th>RSM</th>
<th>RCM</th>
<th>SAM</th>
<th>SSM</th>
<th>SCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10R</td>
<td><img src="image31.png" alt="Image" /></td>
<td><img src="image32.png" alt="Image" /></td>
<td><img src="image33.png" alt="Image" /></td>
<td><img src="image34.png" alt="Image" /></td>
<td><img src="image35.png" alt="Image" /></td>
<td><img src="image36.png" alt="Image" /></td>
</tr>
<tr>
<td>1</td>
<td>7.5YR</td>
<td><img src="image37.png" alt="Image" /></td>
<td><img src="image38.png" alt="Image" /></td>
<td><img src="image39.png" alt="Image" /></td>
<td><img src="image40.png" alt="Image" /></td>
<td><img src="image41.png" alt="Image" /></td>
<td><img src="image42.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td>10R</td>
<td><img src="image43.png" alt="Image" /></td>
<td><img src="image44.png" alt="Image" /></td>
<td><img src="image45.png" alt="Image" /></td>
<td><img src="image46.png" alt="Image" /></td>
<td><img src="image47.png" alt="Image" /></td>
<td><img src="image48.png" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td>7.5YR</td>
<td><img src="image49.png" alt="Image" /></td>
<td><img src="image50.png" alt="Image" /></td>
<td><img src="image51.png" alt="Image" /></td>
<td><img src="image52.png" alt="Image" /></td>
<td><img src="image53.png" alt="Image" /></td>
<td><img src="image54.png" alt="Image" /></td>
</tr>
<tr>
<td>3</td>
<td>Test Case #1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>3</td>
<td>Test Case #9</td>
<td><img src="image55.png" alt="Image" /></td>
<td><img src="image56.png" alt="Image" /></td>
<td><img src="image57.png" alt="Image" /></td>
<td><img src="image58.png" alt="Image" /></td>
<td><img src="image59.png" alt="Image" /></td>
<td><img src="image60.png" alt="Image" /></td>
</tr>
</tbody>
</table>
ARCA

Publications


- **COPYRIGHT ISSUED ON THE SW**
FINAL DISCUSSION, REMARKS AND FUTURE WORKS
Conclusion

▪ **Image**
  - Improvements to CBIR: Telecom-CDVS descriptor, Back-Projection Verification, Query Expansion, and **Heatmap**.
  - In **The Social Picture (TSP)** an huge amount of crowdsourced social images can be collected and explored. It represents a **well suited enabling technology for LTE networks**.
  - **Social-Weigthed-Density (SWD)** maps have been used for reach a novel definition of a saliency model we that we named **Social Saliency**.

▪ **Video**
  - **RECfusion** is a framework designed for automatic video curation driven by the popularity of the scenes acquired by multiple devices.

▪ **3D Data**
  - **Digital Archaeology**: we shown how **3D scanning** and **web sharing** can contribute to the improvement of museum policies in the field of public outreach
  - **Medical Context**: we were able to **represent 3D data with just 2 parameters**, gaining a **compact descriptor** easier to be transfered and browsed through hospital networks
THANKS FOR THE ATTENTION…
QUESTIONS?

Filippo LM Milotta
PhD candidate

milotta@dmi.unict.it
milotta@mail.usf.edu