# FaceMatch: Visual Search by Photos of Missing Persons During a Disaster Event

Eugene Borovikov, Szilárd Vajda, Sema Candemir, Girish Lingappa, Sameer Antani, Michael Gill, George Thoma
Communications Engineering Branch, Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD, 20894, USA

## Introduction

We report on our FaceMatching research and development (R&D) that aims to provide robust image near-duplicate detection and face localization/matching on digital photos of variable quality, as an integral part of PEOPLE LOCATOR (PL)® developed by NLM as a Web-based system for family reunification in cases of natural or man-made disasters. PL collects photos and brief text meta-data (name, age, etc.) of missing or found persons. Currently supported text queries may be insufficient because text data are often incomplete or inconsistent. Adding an image search capability can significantly benefit the user experience. Face localization is done via skin-tone/landmarks enhanced gray-scale face detector, more accurate than many open source and commercial detectors. Face matching is done via an ensemble of image descriptors (HAAR, LBPH, SIFT, SURF, ORB), using a smart re-ranking procedure. We describe the integration of our face matching system with PL, report on its performance. Unlike other face recognition systems often having many good quality well-illuminated sample images for each person, ours can handle the lack of training examples for individual faces, as those are unlikely in a disaster setting.

### Challenges

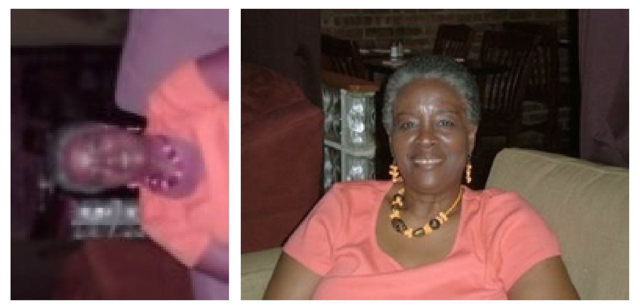
- low quality, suboptimal lighting
- pictures may contain $0$ or more faces
- face-like objects (animal/cartoon faces)
- presence of duplicates and near-duplicates
- face images may be hard to match due to
  - partially occluded or damaged faces
  - presence of facial hair, glasses, jewelry
  - person natural aging
  - source photograph degradation

## Near-Duplicate Detection

### Description

An image data-set may contain many near-duplicate images due to multiple postings of the same photograph rescaled or re-compressed. Such near-duplicates need to be identified and grouped. Each group would be represented by the highest quality image. We solve this by

- color wavelet based descriptor: most significant wavelet coefs'
- real-valued distance measure in $[0, 1]$, with $0 =$ perfect match
- tight threshold for *near*-duplicate detection
- champion selection: highest resolution, lower compression
- using 128x128 YIQ color images: gray-scale compatible
- being robust to scale and re-compression



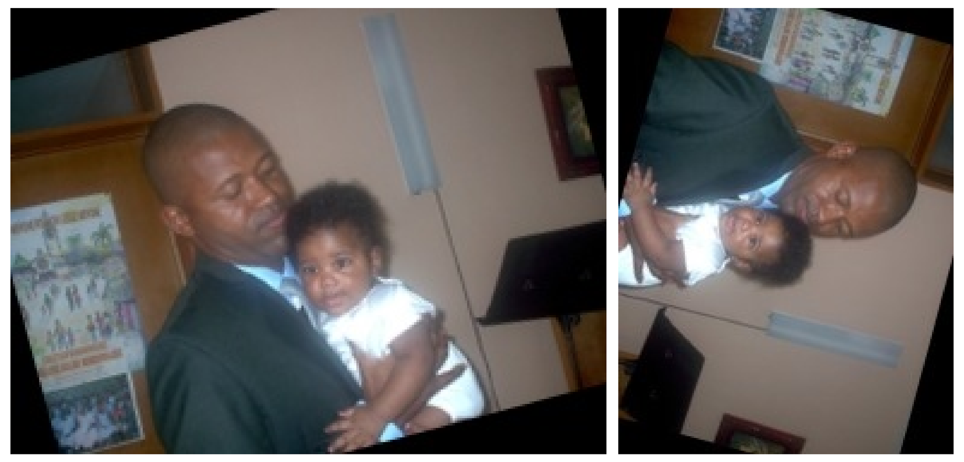
### Experiments

Detect near-duplicate images in our data

| data-set | | near-duplicates | | |
|---|---|---|---|---|
| name | size | # of | % of | proc.time |
| HEPL | 15K | 6K | 40 | 5 min |
| PL | 12K | 4K | 30 | 4 min |

Image matching on generated near-dups

| distortion | Recall | Precision | F-score |
|---|---|---|---|
| rotation | 0.69 | 0.62 | 0.65 |
| crop | 0.71 | 0.70 | 0.71 |
| scale | 0.99 | 0.99 | 0.99 |



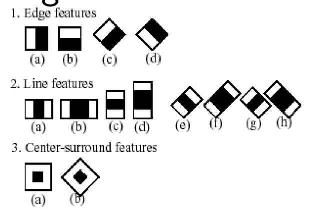
We have also experimented with generating 792 near-duplicates from a set of 132 unique images by scaling ($s = 0.5, 2$), rotating ($\alpha = \pm\pi/12$) and cropping ($c = 0.8, 0.65$). Our near-duplicate detector is most sensitive to rotations and cropping, detecting very few of those, while detecting most of the scaled near-duplicates correctly. This result was rather expected, given the Haar wavelet nature of the detector.

## Face Detection

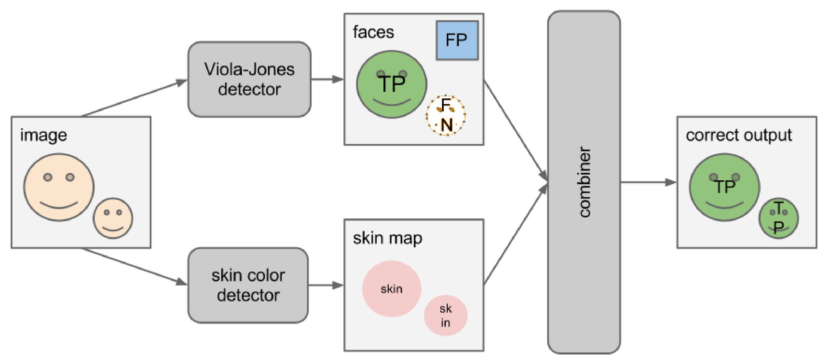
### Description

A reliable *face detector* is necessary for any face matching application, as it determines the locations and sizes of human faces in digital images. Our FaceFinder achieves this goal via

- Haar-like gray-scale features
- major 90-degree rotations
- skin color mapping in RGB, HSV, Lab spaces
- color based landmarks (eye, nose, mouth) detection
- artificial neural net (ANN) landmark verifier
- correcting minor rotations using eye line

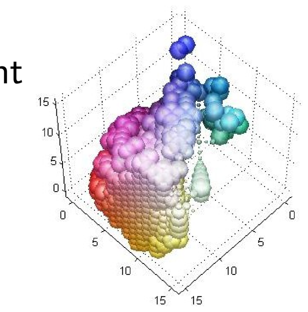


### ViolaJones+SkinMap+Landmarks



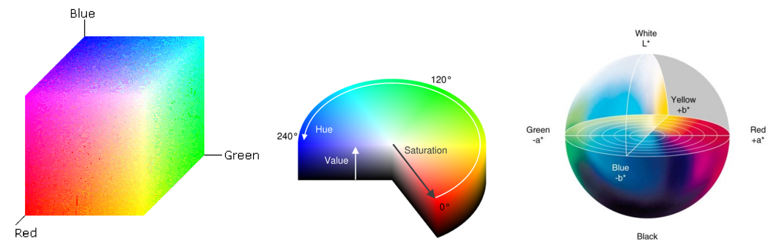

### combiner module

- recover false negatives (FN)
  - skinmap driven enhancement
  - color landmark detection
- reject false positives (FP)
  - skinmap region integration
  - landmark positioning

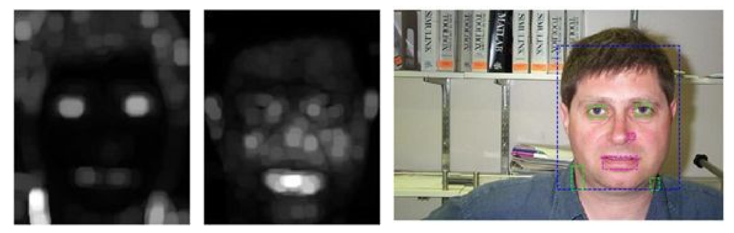


### skin mapping

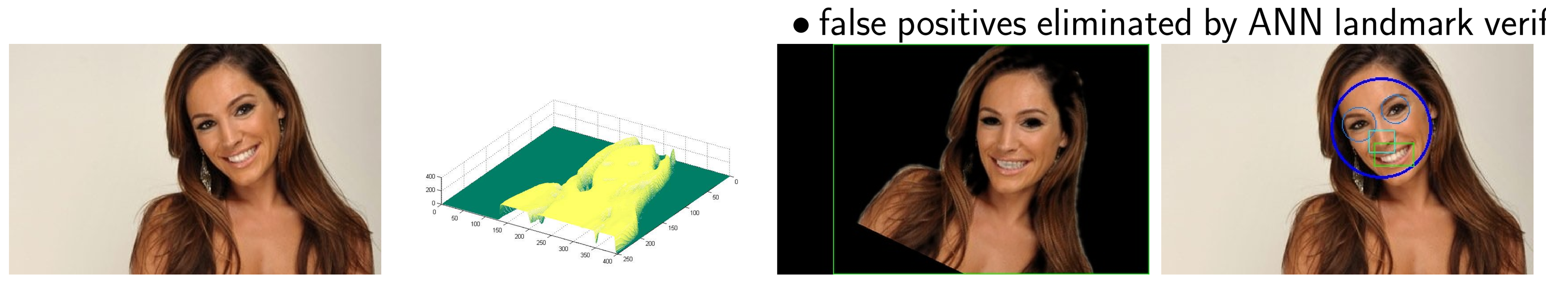
- skin color information from annotation
- estimating color models: ANN, histogram
- extended color space: [RGB,HSV,Lab]



### landmark detection



- eye/mouth maps derived from luma/chroma bands
- major peaks are landmark candidates
- false positives eliminated by ANN landmark verifier



### Experiments

With no modifications, Viola-Jones face detector misses about half of the PL faces. About 20% of these are typically too small for matching. The data-sets we experimented with:

**HEPL-500**: 500 images from Haiti

**Lehigh-512**: 512 celebrities images

**Caltech-450**: 450 Caltech faculty faces

Aided by skin mapping and landmark awareness, our FaceFinder outperforms some major commercial detectors (iOS, FaceSDK) and the leading open-source detectors by Viola-Jones and Zhu-Ramanan.

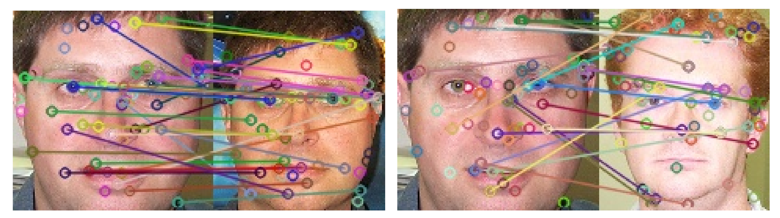
| data-set | method | R | P | F |
|---|---|---|---|---|
| HEPL-500 | ViolaJones | 0.76 | 0.87 | 0.81 |
| | **FaceFinder** | **0.77** | 0.89 | **0.83** |
| | iOS | 0.68 | 0.87 | 0.76 |
| | FaceSDK | 0.73 | 0.87 | 0.79 |
| | Zhu-Ramanan | 0.33 | **0.92** | 0.49 |
| Lehigh-512 | ViolaJones | **0.95** | 0.81 | 0.88 |
| | **FaceFinder** | **0.95** | **0.94** | **0.94** |
| | iOS | **0.95** | 0.92 | **0.94** |
| | FaceSDK | 0.93 | 0.91 | 0.92 |
| | Zhu-Ramanan | 0.83 | 0.91 | 0.87 |
| Caltech-450 | ViolaJones | 0.95 | 0.88 | 0.91 |
| | **FaceFinder** | **0.98** | 0.97 | **0.98** |
| | iOS | 0.97 | **0.98** | 0.97 |
| | FaceSDK | 0.96 | 0.94 | 0.95 |
| | Zhu-Ramanan | 0.97 | 0.97 | 0.97 |

## Face Matching

Once the face/profile regions in the image collection are localized and their descriptors are indexed, they can be matched against a query face/profile picture, which may come from an existing (possibly annotated) image, or from a new photograph, that FaceMatcher has not seen before. Hence the face matching method needs to be robust to accommodate wide variations in the appearance, and it needs to be fairly exact to eliminate numerous false positive hits.

### Solution

- localized face/profile
- HAAR/SIFT/SURF/ORB descriptors
- scale/rotation invariant metrics
- distance range $[0, 1]$
  - $0 =$ perfect match
  - $1 =$ complete mismatch
- ensemble approach to *trainingless matching*



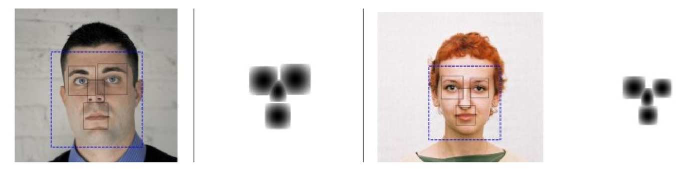
### Improvements

- candidate list re-ranking based on

  **MANY**: $d = \sqrt{d_1\sqrt{d_2 \ldots \sqrt{d_n}}}$ with $d_1$ being the most confident (heaviest) distance

  **DIST**: $d = \prod d_i^{w_i}$ with the constituent distances and their weights typically in $[0, 1]$

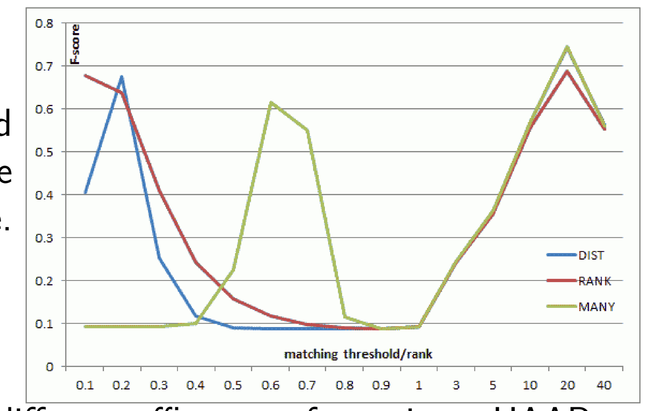
  **RANK**: Borda count weighted re-ranking
- stronger descriptors weigh more
- downplay weak matches via salience maps



### Experiments

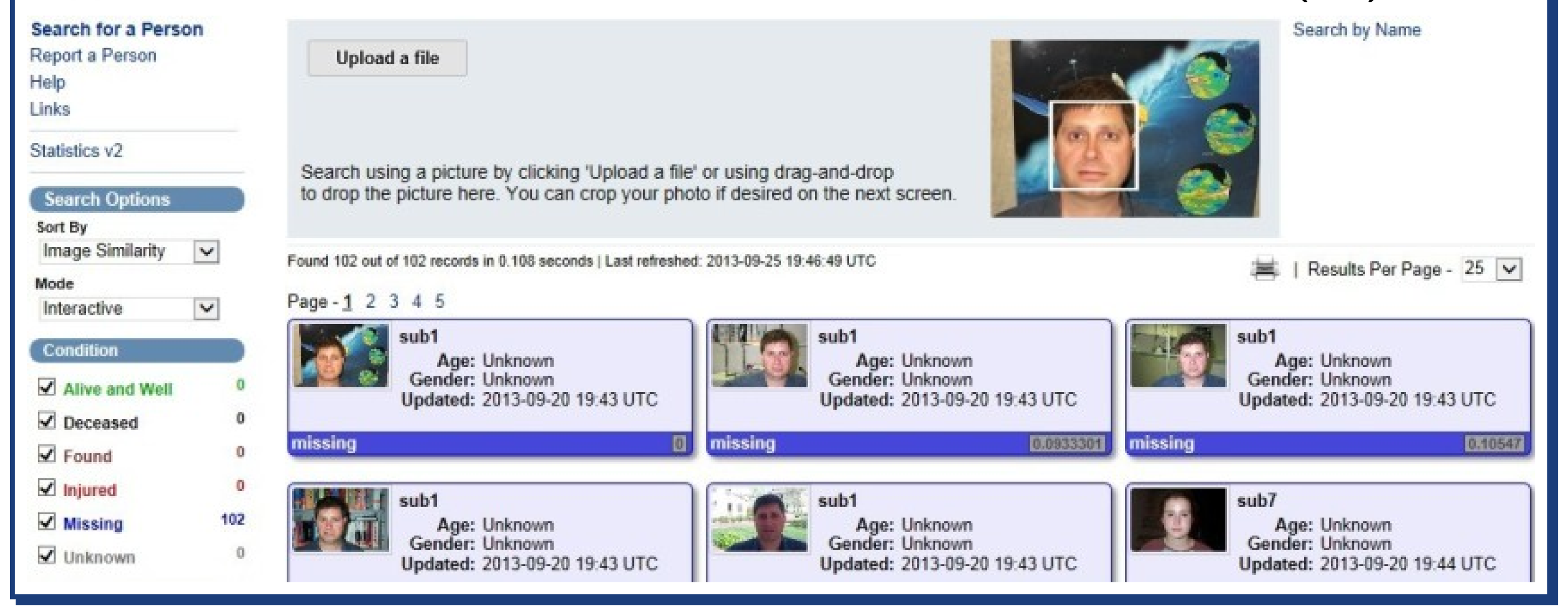
We experimented with Caltech (450 color images) dataset and HEPL-372 (62 images with 6 synthetic modifications: crop, scale and rotate). Accuracy (F-score) figures are reported in the table.

| method | HAAR | SIFT | SURF | ORB | MANY | DIST | RANK |
|---|---|---|---|---|---|---|---|
| HEPL-372 | 0.67 | **0.91** | 0.88 | 0.71 | **0.97** | 0.96 | 0.86 |
| CalTech | 0.25 | **0.64** | 0.61 | 0.58 | **0.74** | **0.74** | 0.69 |




SIFT descriptor (slowest to compute) was the most robust to different affine transformations. HAAR was the fastest, but the least accurate. ORB was also fast, but not as accurate as SIFT or SURF. A weighted ensemble was always more accurate than any individual descriptor.

### Application: FaceMatch web services for PEOPLE LOCATOR (PL)®




## Conclusion

We provide query-by-image capability to the PEOPLE LOCATOR (PL)® system, evaluated a few state-of-the-art systems on existing data-sets and developed tools for image annotation and near-duplicate detection. The face detection module improves a gray-scale face detector with the skin/landmark detection techniques. The face matching subsystem uses an ensemble of descriptors to capitalize on the strengths of its constituents, and results in higher accuracy figures than any of the individual descriptors.

27th Annual NIH Research Festival
Bethesda, MD, USA October 7-11, 2013

CEB