NATIONAL LIBRARY OF MEDICINE

NATIONAL INSTITUTES OF HEALTH

FaceMatch: Visual Search for Pictures of Missing Persons During a Disaster Event

Eugene Borovikov, Szilárd Vajda, Girish Lingappa, Sameer Antani, Michael Gill, George R. Thoma

Communications Engineering Branch, Lister Hill National Center for Biomedical Communications, U.S. National Library of Medicine, Bethesda, MD, 20894, USA

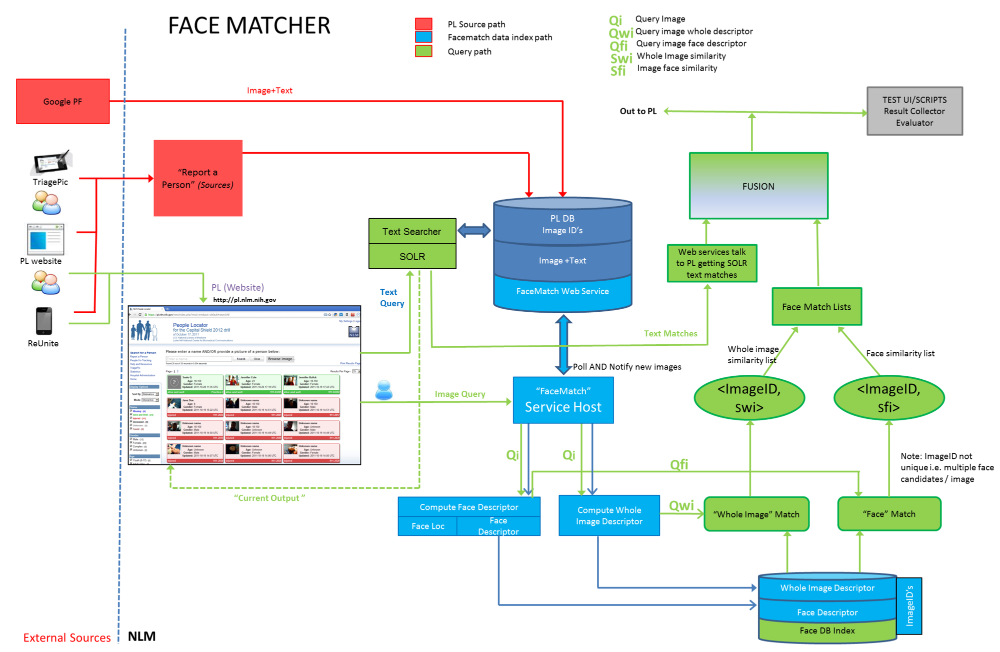
Abstract

NLM's People Locator™(PL) system allows the posting of photos and simple metadata (name, age, location) for persons missing (or found) in the wake of a disaster. To extend the current text-based search method with a visual search for people's faces, we developed FaceMatch, a system to match faces in a query image to those in the stored photos. Face matching is a two-stage process: faces in photos sent as queries are first localized using an improved Viola-Jones face detector, and then image features (SIFT, SURF, ORB and HAAR) are extracted, combined and matched against an index of features extracted from the stored photos. Face matching in this context is challenging because of the lack of training data, low-resolution photos, wide variability in lighting, facial expression, head pose, ethnicity, occlusions and deformed faces due to injury. Ongoing research includes exploring more discriminating features, modeling skin color for more accurate face localization, a Haar wavelet-based technique to eliminate near-duplicate photos, and image normalization. The approach is tested on images collected from the 2010 earthquake in Haiti (HEPL collection) and Labeled Faces in the Wild (LFW dataset). Current FaceMatch speed and accuracy performance results are presented.

Overall Challenges

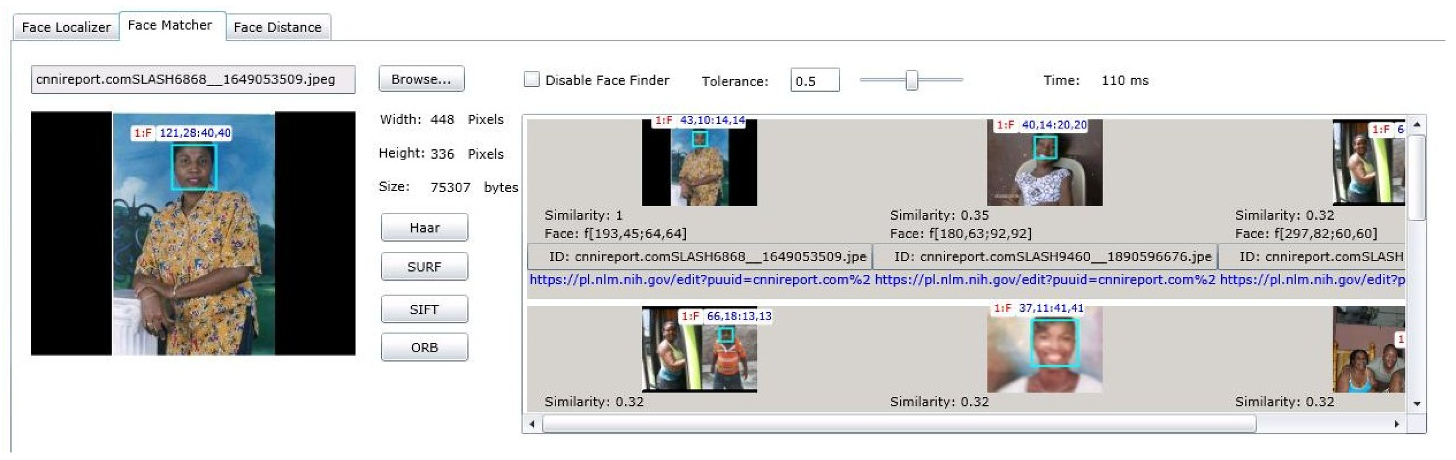
- pictures may contain 0 or more faces
- face-like objects (cats and dogs faces)
- query/database images may be of suboptimal quality due to:
  - partially occluded or damaged faces
  - presence of duplicates and near-duplicates
  - inconsistency due to facial hair, glasses, jewelry, aging

System Overview



Application

FaceMatch web services integrate in PL as an intuitive and accessible complement to text search.



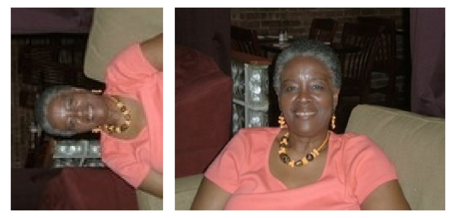
Near-Duplicate Detection

Description

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

Solution

- Haar wavelet based descriptor: most significant wavelet coefs'
- real-valued distance measure in $[0,1]$, with 0 = perfect match
- low threshold for *near*-duplicate detection
- champion selection: highest resolution



Near-Duplicate Experiments

Number of near-duplicates in data-sets

- HEPL: 6K near-dups in 15K images
- PL: 4K near-dups in 12K images

We have also experimented with generating about 800 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.

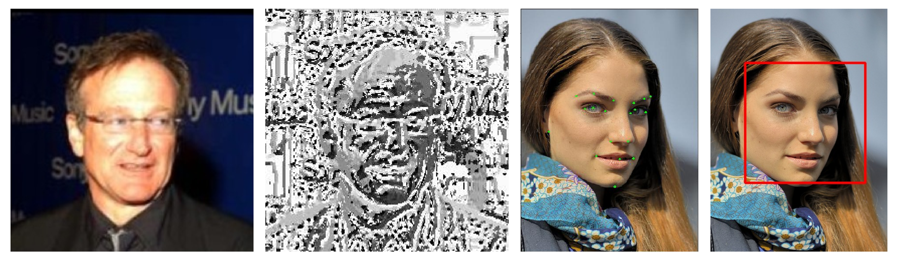
Face Detection

Description

Face detection determines the location and the size of a human face in an arbitrary digital image using:

- low-level image features (Haar, LBP[a])
- high-level facial landmarks: eye, nose, mouth, ear, chin, etc.
- skin color

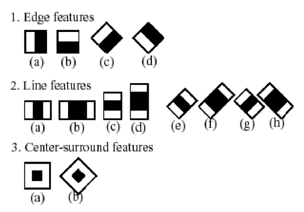
[a] T. Ojala, M. Pietikinen, and D. Harwood, Performance evaluation of texture measures with classification based on Kullback discrimination of distributions, ICPR, 1994



Solution

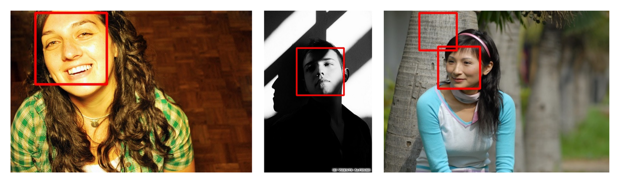
Method

- Haar like features
- multi-scale window technique
- Ada boost classifier cascade (Viola-Jones[a])



Drawbacks

- small size faces not detectable
- lighting/occlusion deteriorates results



[a] P. Viola and M. Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, CVPR, 2001

Improvements

- color information (skin)
- learning color models
- extended color space + neural network

Advantages

- high accuracy for skin detection (91%)
- skin maps focus the face finding
- enhance skin region intensities

Face Detection Experiments

With no modifications, Viola-Jones face detector misses about half of the HEPL faces, and about 20% of the missed ones are typically too small for matching.

Accuracy results of our FaceFinder

- HEPL-300 (300 suitable images): R = 75%, P = 83%, F = 79%
- HEPL-4K (4,000 images including noise): F = 53%
- PL-700nd (700 faces skipped by Viola-Jones) + skin map: 46% Recall boost, 22% False Positives rate

The major factors hurting the accuracy are: Lighting = 5.3%; Low Quality = 8.8%; Occlusion = 10.8%; Color = 0.6%; Combination = 9.5%; Small Faces = 20.8%; Other = 44.2%.

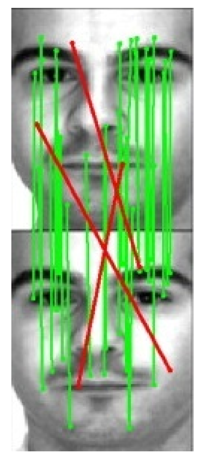
Face Matching

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

Method

- localized face/profile
- Haar/SIFT/SURF/ORB descriptors
- scale invariant metrics
- distance range $[0,1]$
  - 0 = perfect match
  - 1 = complete mismatch



Keypoints correspondence

Improvements

- re-ranking based on
  - distance average
  - Borda count
- stronger descriptors weigh more
- downplay weak matches

Face Matching Experiments

Our experiments with the annotated HEPL-4K images, and on HEPL-62mod (372 = 62 images with 6 synthetic modifications, e.g. crop, scale and rotate). Accuracy (F-score) figures are reported in the table.

Dataset	HAAR	SIFT	SURF	ORB
HEPL-4K	**0.99**	0.98	0.96	0.95
HEPL-62mod	0.44	**0.81**	0.52	0.56

We have also experimented with combining the descriptor match distances by using a generalized geometric mean, and found that a combination of HAAR*ORB*SIFT*SURF produce a better F-score (by about 5%) than either of the individual descriptors. Inclusion of a weak descriptor tends to hurt the ensemble.

Conclusion

Having a goal to enable image based query capability in the People Locator™(PL) system, we studied several image matching and face recognition methods, evaluated a few state-of-the-art systems on existing data-sets and developed core tools for image near-duplicate detection, face detection and face matching. The near-duplicate image detector tool [Jacobs et al.,1995] helps the DB administrator to clean up or group near-duplicate images in the PL data-base. The face detection capability relies on Viola-Jones object detection method, improved by the skin detection techniques. The face matching subsystem uses Haar, SIFT, SURF and ORB descriptors in an ensemble to capitalize on the strengths of its constituents, and results in higher accuracy figures than any of the individual descriptors.

26^{th} Annual NIH Research Festival
Bethesda, MD, USA, 2012 October 9-12

CEB