SiblingsDB – Content description

The SiblingsDB contains different datasets depicting images of individuals realted by sibling relationships. Actually, the images are organized in two different DBs.

The first, called HQ_{faces}, contains a set of high quality images depicting 184 individuals (92 pairs of siblings). A subset of 79 pairs contains profile images as well, and 56 of them have also smiling frontal and profile pictures. The images, with resolution 4256x2832, were shot by a professional photographer with uniform background and controlled lighting. The subjects are voluntary students and employees of the Politecnico di Torino and their siblings, in the age range between 13 and 50 (average 23.1). The highest and average age differences between siblings are 30 and 4.6 years, respectively. All subjects are Caucasian and 57% of them are male. Subjects were asked not to wear make-up. The images have been organized into three Individual Datasets (IDS), namely *HQf*, *HQfp* and *HQfps*:

- *HQf*: frontal expressionless images of 184 subjects (92 sibling pairs);
- *HQfp*: 158 individuals, each represented by one frontal and one profile expressionless images (79 sibling pairs);
- *HQfps*: 112 individuals, each represented by a set of four images per individual. Two expressionless frontal and profile, and two smiling frontal and profile images (56 sibling pairs).

All the images are annotated with, respectively, the position of 76 landmarks on frontal images and 12 landmarks on profile images. The indices of the landmarks in the images are shown in Fig. 1. For each individual the information on sex, birth date, age and votes of the panel of human raters (who were asked to evaluate if the couples depict siblings or not) are also available. For each IDS, we created a Pair Dataset (PDS) for our experiments containing all the positive pairs, the N/2 pairs of siblings, and an equal number of randomly chosen negative, non-sibling, samples. For simplicity, in the following the PDS have the same name as the IDS used to build them (*HQf*, *HQfp* and *HQfps*).

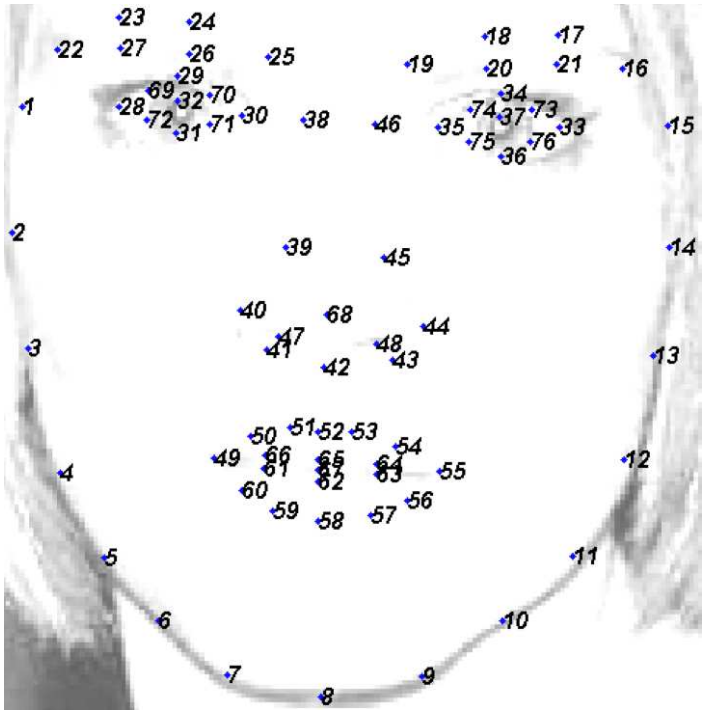

Fig. 1. Landmark positions

Data are organized as follows:

- subjects.csv includes the individual information: ID, Sex ("M" or "F"), birth date, age when the pictures were taken (in 2010). N.b.: not all subject images are available, as some of them where not included in the DB
- HQxx.csv: a file for each PDS containing a line for each couple used in the experiments. This line includes the following information
    - unique couple ID: IDs of the individuals in the couple in the form ID1_ID2
    - are they siblings?: 1 if the two individuals are siblings, -1 if not
    - # votes: total number of votes from the human panel of raters (HP)
    - Yes' answers: number of HP yes answers
    - No' answers: number of HP no answers
    - first picture: list of the pictures depicting each individual (a different number of images are available for each individual according to the IDS: a frontal expressionless image for *HQf*, a frontal and profile expressionless images for *HQfp*, frontal and profile both expressionless and smiling images for *HQfps*). The list includes the path to each image, coded as *individualDIR/image.jpg*. Images are stored in '*DBs'* directory into a subdirectory corresponding to the IDS (one of *HQf*, *HQfp* and *HQfps*); inside them, *individualDIR* is the subdirectory containing the images for each individual and *image* is the image filename. When an individual is represented by more than one image, the image filenames are separated with the '+' character (e.g. filename1+filename2+filename3+filename4 for individuals in *HQfps*). The file containing the position of the landmarks for each image are included in a file named *individualDIR/image.cvs*
    - second picture: same as before
- HQxx_sibling_matrix.mat: a file for each PDF containing a matrix where the element (i,j) is 1 if samples i and j are siblings
- individualDIR/image.csv: a file for each image containing the list of the landmark positions (x,y)

The second dataset is called LQ_{faces}, and includes a set of low quality images downloaded from the Internet. All samples are included into the LQf/raw wich also contains a csv file for each image with the list of the landmark positions (x,y). The filename are in the format *surname_name.jpg*, from which it is immediate to obtain the sibling pairs. The non-sibling pairs are obtained circularly shifting the second brothers of one position. In other words, imagine to sort alphabetically all samples. Then positive pairs are 1-2, 3-4, 5-6, 7-8 and so on. Negative samples are 1-198, 3-2, …, 197-2.

Experimental protocol

Two different experiments were performed in our article:

1. within DB experiments. Each PDS (*HQf*, *HQfp, HQfps* and LQ_{faces}) was classified assessing results using stratified five-fold cross-validation (CV); the accuracies reported in the paper are the average classification rates of the classifier over the different CV rounds.
2. Cross-DB experiment. HQf PDS was used for training and LQ_{faces} for testing

Please refer to the following paper for further details on the DB and on the experimental results:

T.F. Vieira, A. Bottino, A. Laurentini, M. De Simone, Detecting Siblings in Image Pairs, The Visual Computer, 2014, vol 30, issue 12, p. 1333-1345, doi: 10.1007/s00371-013-0884-3