Automatic face labeling system for the National Archives of Estonia

The National Archives of Estonia has around 500,000 photos in their digital archive. To enhance the search capabilities of the archive, it would be nice if people on those photos would be named or labeled. Unfortunately manually labeling hundreds of thousands of images is not feasible, so automatic labeling system was proposed in Jan van Gent’s bachelor thesis “Clustering Faces from the National Archives of Estonia”. The task for this project is to use ideas and code from this thesis to implement a prototype web system to collect crowd-sourced face labels from general public. The final goal for the system is to be integrated with Fotis digital photo database.

The system should include following functionality:

- Automatic detection of faces from entire database (500,000 images)
- Grouping of face images based on similarity (code available from the thesis)
- List of groups (ordered by number of images, i.e. bigger groups first)
- List of images in a group
- Searching group by name
- Searching image by description
- Labeling a group (assigning a name to it)
- Labeling a single image (assigning it to a different group)
- Suggestion of labels to a group (based on similar groups, i.e. merging of groups)
- Suggestion of labels to an image (based on similar images, i.e. change group)
- Automatic reassignment of groups based on similarity (manually assigned labels cannot be overridden by automatic process, automatically assigned labels can)
- Manual confirmation of labels assigned by the automatic system
- Automatic evaluation of crowd-sourced labels, i.e. need 5 people to confirm the label before it is fixed, or the label must come from a trusted user.

Many user interface hints can be taken from Google Picasa application. While it is officially not supported any more, it had a convenient interface for labeling large number of photos. Older versions can still be found on the web.

The system should use database structure that can be later merged with Fotis database at the National Archives.

Contacts:
Tambet Matiisen, tambet.matiisen@ut.ee
Lauri Leht, lauri.leht@ra.ee