Image fragment matching in practice: applications and supplementary tools

Andrzej Śluzek

Nanyang Technological University
Singapore

Nicolaus Copernicus University
Toruń

Acknowledgments

The presented work has been developed within, and is a part of A*STAR Science & Engineering Research Council (Singapore) grant 072 134 0052 **"Framework for Visual Information Retrieval and Building Content-based Visual Search Engines"**.

The financial support of SERC is gratefully acknowledged.

Overview

- Automatic formation of „visual objects" in unknown worlds

  Visual prototypes : *nodes of similarity graphs*

  Visual objects : *sub-graphs of similarity graphs*

- Analysis of face similarities based on image fragment matching

  Fragment matching as a formal model of subjective face similarities

  Simultaneous face detection and recognition

  Databases of face images – analysis and optimization

- Image fragment matching for speed control in adverse conditions

  Size and continuity of similar fragments as a model of visibility conditions

  Speed control based on visibility conditions

  Real-time tracking of similar fragments in video streams

- Supplementary techniques

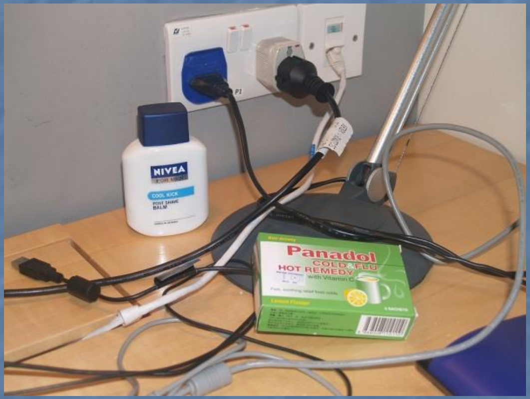
  Alternative keypoint descriptor of lower dimensionality

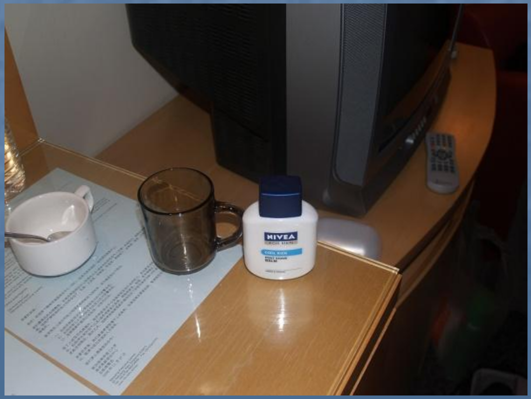
  Image co-segmentation based on TPS warping

Automatic formation of „visual objects" in unknown worlds

Visual prototypes

A small database of images





Exemplary pairs of near-duplicate fragments

Topological approach can be used, but the affine method is preferred because it generates near-duplicates with less contents of backgrounds.

Visual prototype: a cluster of overlapping near-duplicates from the same image (each near-duplicate found by matching to a different image from the database)

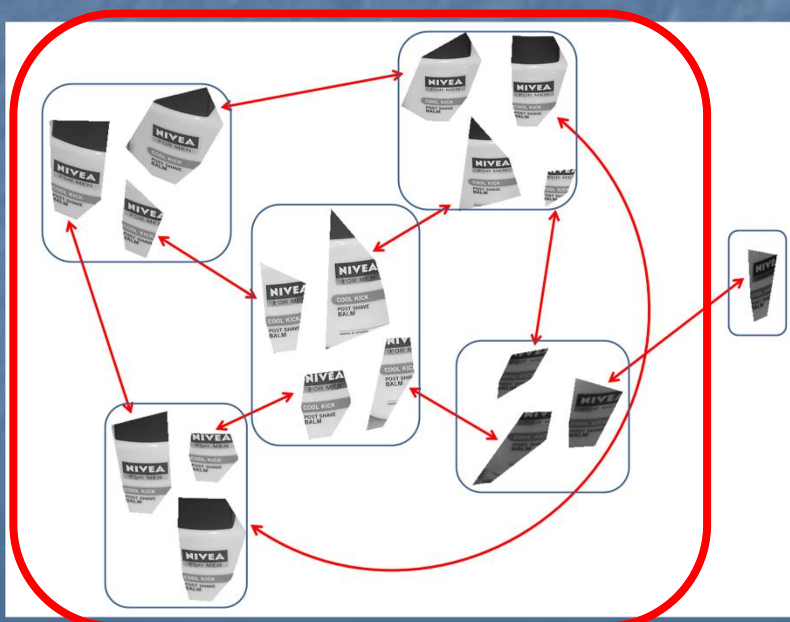
$$\min\left(\frac{|f_1|}{|f_2|},\frac{|f_2|}{|f_1|}\right) > Tr_1$$

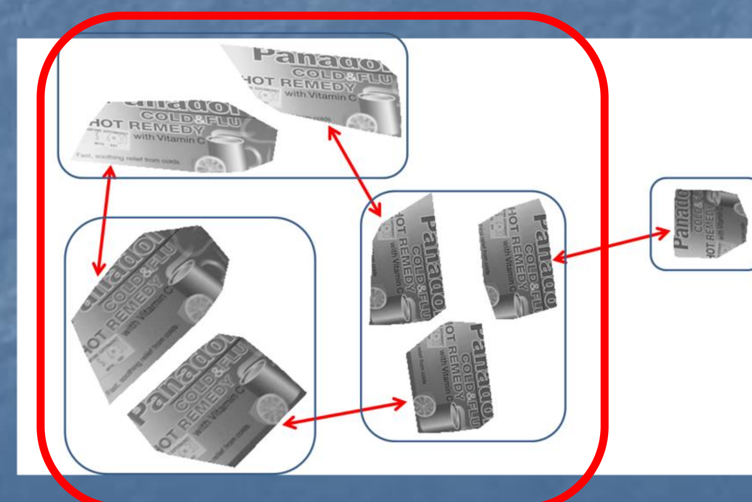
$$\frac{|f_1 \cap f_2|}{\min\left(|f_1|,|f_2|\right)} > Tr_2$$

Visual objects

Similarity graph: each node is a visual prototype; edges are formed naturally by linking the corresponding near-duplicates.

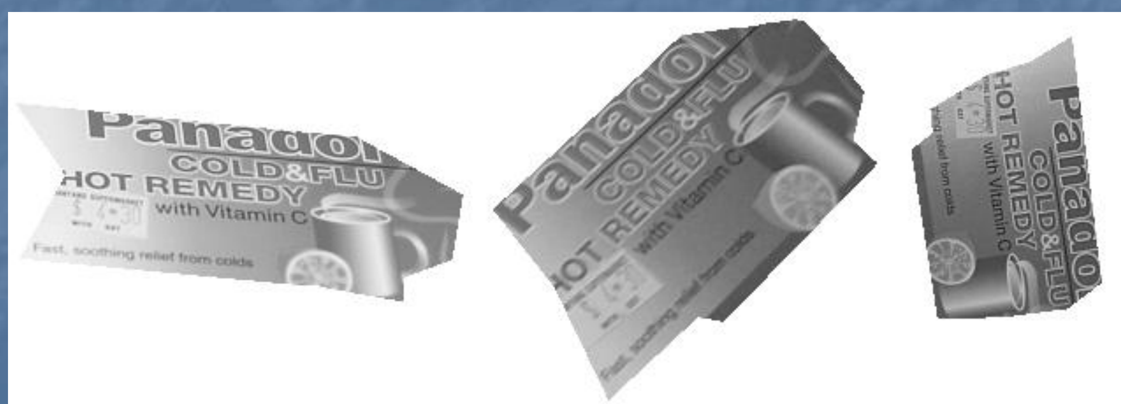
VISUAL OBJECTS are defined as as **2-connected sub-graphs** of the similarity graph. In other words, a visual prototype must be similar to at least two other visual prototypes to be included into a visual object.





Once a *visual prototype* is includes into a *visual object*, the union of all near-duplicates within this prototype forms the instance (template) of this *visual object* within the corresponding images.

Two visual objects found in the exemplary database. Their instances are:

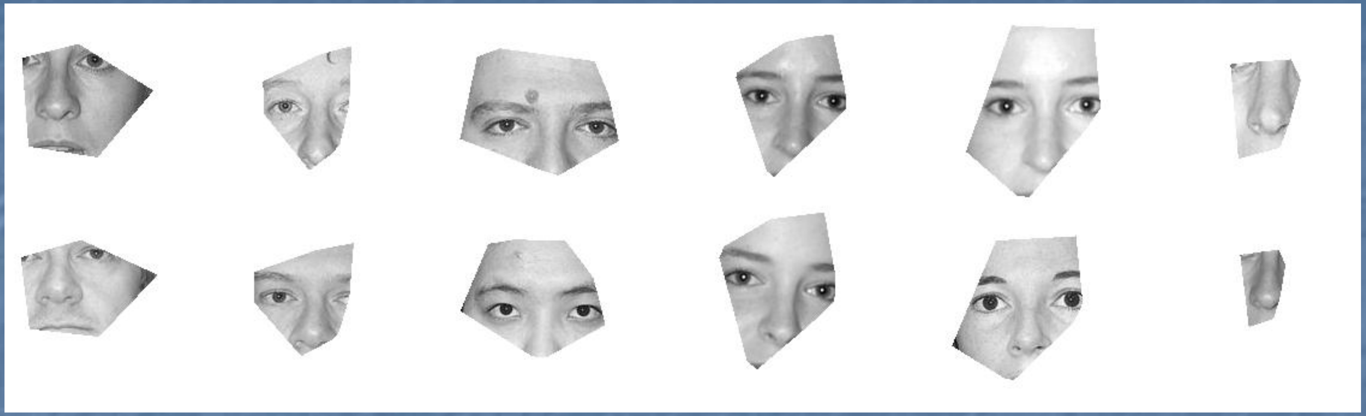


Analysis of face similarities based on image fragment matching

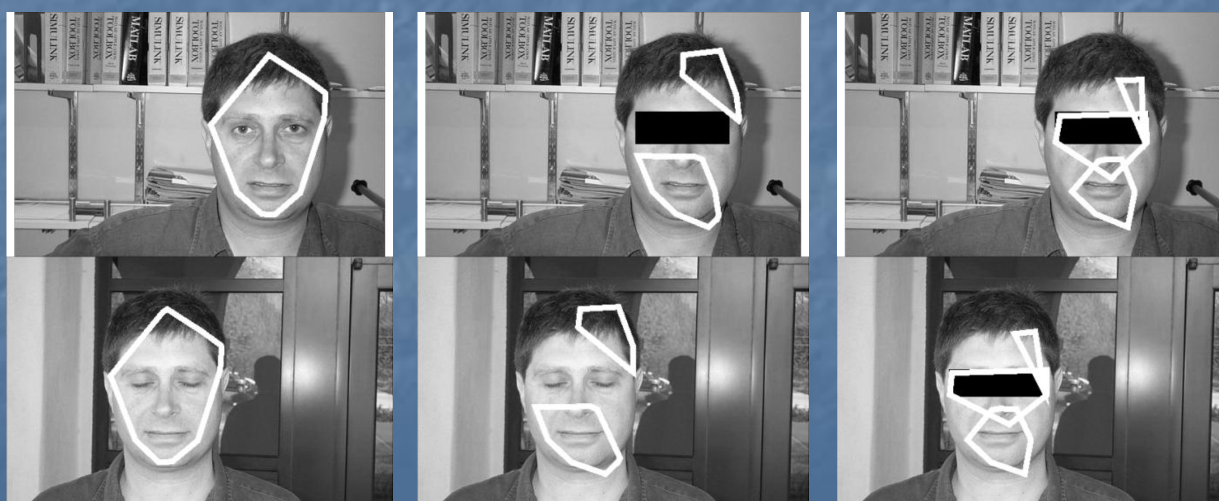
Statements like: "*he looks so different at this photograph*", "*they have so similar eyes and mouths*", "*moustache has completely changed his face*", "*at these photos they look like twins*", etc. sound subjective but they apparently represent some objectively existing or missing visual similarities between images of face (and, therefore, the face identification decisions may be affected).

Topological approach can be prospectively used, but most faces are topologically similar so that the geometric approach is again preferred .

We propose to use **image fragment matching as a formal model of subjective face similarities**



The approach is, for example, highly insensitive to standard methods of anonymization.



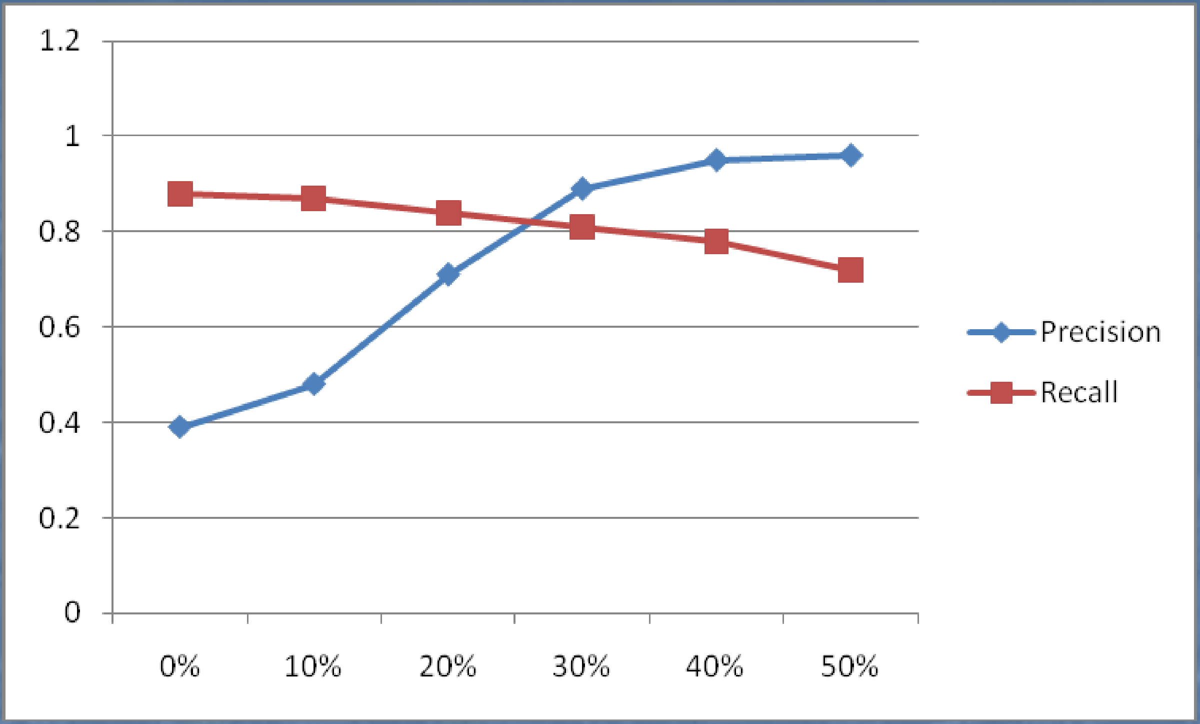
Where is the difference between **detection of similar faces (face fragments)** and **face recognition**?

Pre-requisite: Faces of known people are outlined in database images.

The overlap between the database face outline and the detected near-duplicate (the face coverage) plays the crucial role.

A high coverage requirement provides more credible identification of the person. By accepting lower coverages we increase chances of a correct identification for differently looking photos of the same person, but also increase the risk of mistake.

Exemplary results for Caltech Faces (but not Easy Faces)



For lower coverage, most *the-same-face* pairs are but more *false positive* mistakes exist as well. At 40% coverage, however, only 5% of returned near-duplicates are *false positives*, but *there are more cases of missed.* Even for 0% coverage 12% of *the-same-face* pairs are not identified. In our opinion, such pairs represent cases which can be subjectively described by statements: "*he looks so different at these photos*", "*I can hardly believe this is her*", etc.

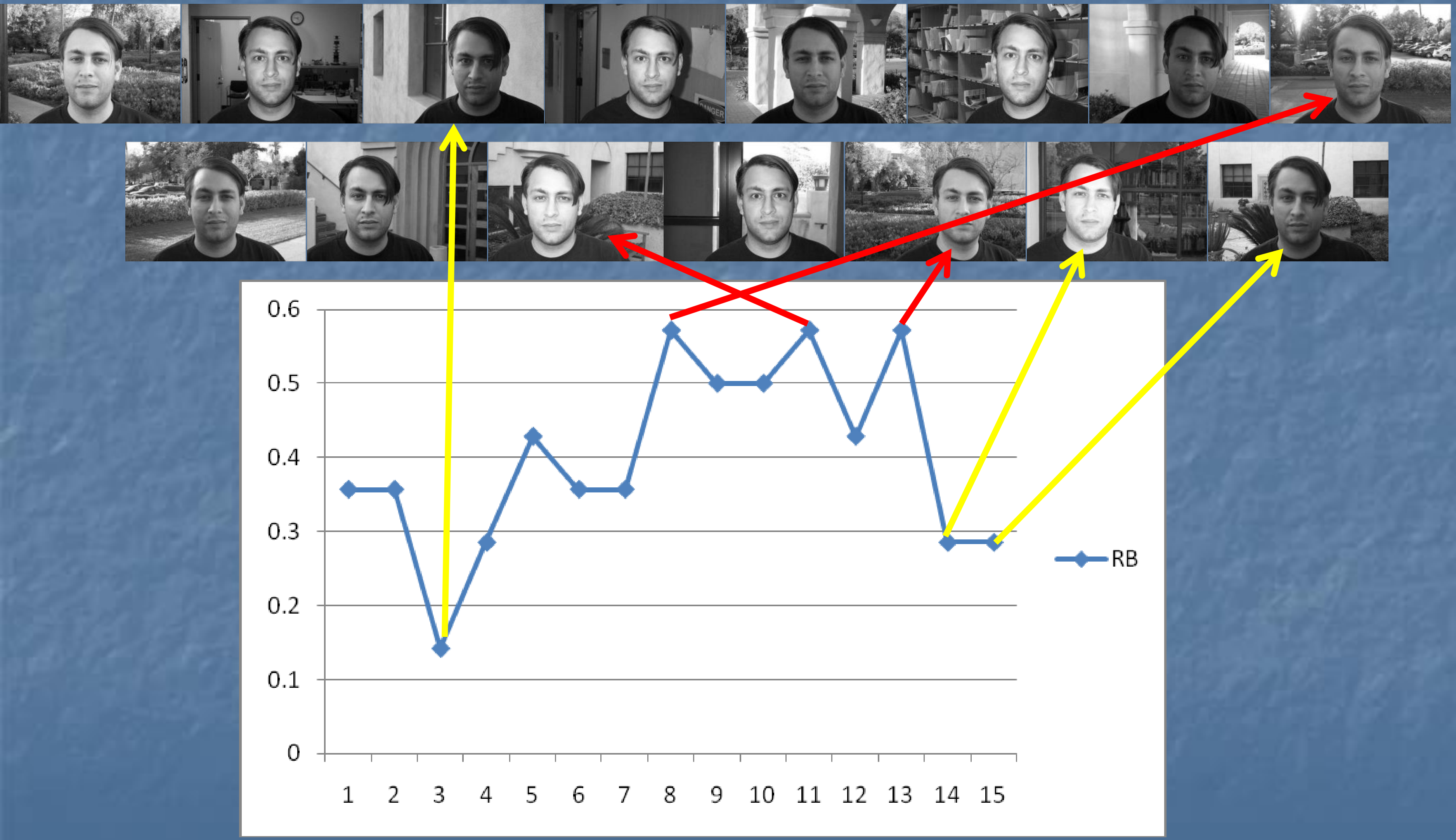
Databases of face images – analysis and optimization

Two database images containing a human face are considered similar if a near-duplicate intersecting the face outlines is detected in both images. This will be referred to a *binary* measure of similarity S_B. A *continuous-range* measure of similarity S_C between two face images can be obtained by using the average *coverage* of face outlines by the near-duplicates in both images.

Given the similarity measures S_B and S_C, and a collection of n images presenting the same face, the relevance (representativeness) of an individual image I_i from this collection can be expressed as

$$R_B(I_i) = \frac{1}{n-1}\sum_{\substack{k=1 \\ k\neq i}}^{n} S_B(I_i, I_k) \qquad R_C(I_i) = \frac{1}{n-1}\sum_{\substack{k=1 \\ k\neq i}}^{n} S_C(I_i, I_k)$$

The most representative face images



Inter-class similarity of face photos

Given two classes of face images $\boldsymbol{C_I}$ (containing I_1, I_2,..., I_m images) and $\boldsymbol{C_J}$ (containing J_1, J_2,..., J_n images) the inter-class similarity *SC* can be straightforwardly defined as

$$SC_B(C_I, C_J) = \frac{1}{mn}\sum_{k=1}^{m}\sum_{l=1}^{n} S_B(I_k, J_l) \qquad SC_C(C_I, C_J) = \frac{1}{mn}\sum_{k=1}^{m}\sum_{l=1}^{n} S_C(I_k, J_l)$$

Higher values of *SC* increase the risk that images of $\boldsymbol{C_I}$ faces can be identified as faces from $\boldsymbol{C_J}$ class (or another way around).

Using *SC* measures, we can identify individuals with similar faces, with the most „unique” faces, the most „average” faces, etc.

Examples

The most similar persons in the database of 19 individuals (15 photos for each face)

The most „unique” face.

The most „average” faces.

Additional results

Using the method of *automatic formation of viisual objects*, many images of the same face are clustered into the same „visual object".

Examples of near-duplicates which contribute to the same „visual object":

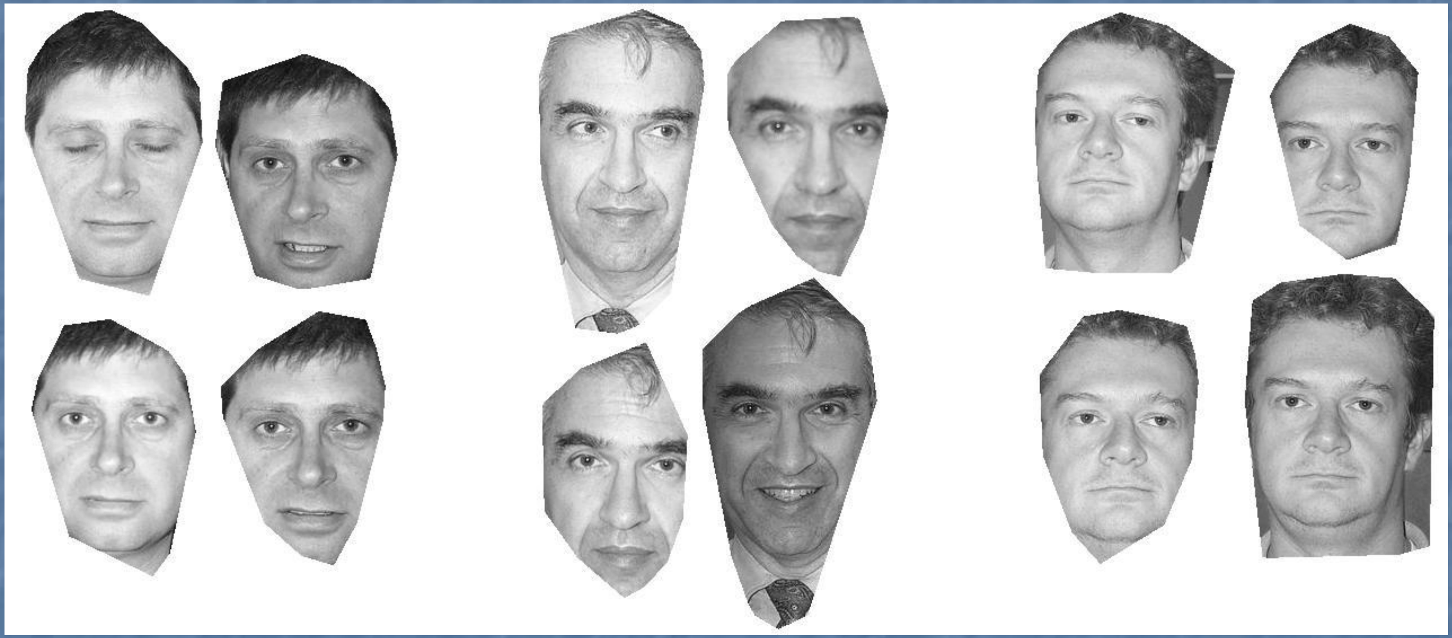


Image fragment matching for speed control in adverse conditions

From the vehicular perspective, good visibility can be defined as conditions when objects and components of the observed moving world can be smoothly and unobtrusively tracked. It can be argued, that matches between contents of frames captured by a video-camera can be a feasible model of the visibility condition.

Frames should be separated by ***T*** frames (for neighbouring frames a good match exists even in difficult conditions).

The proposed model includes:

Defining the region of interest ***R*** in the captured video. Typically, R occupies the central part of frames.

A measure of visibility quality ***VQ***:

$$VQ(N) = \frac{\left\| R \cap D_N^{N-T} \right\|}{\left\| R \right\|}$$

A measure of visibility continuity ***VC***:

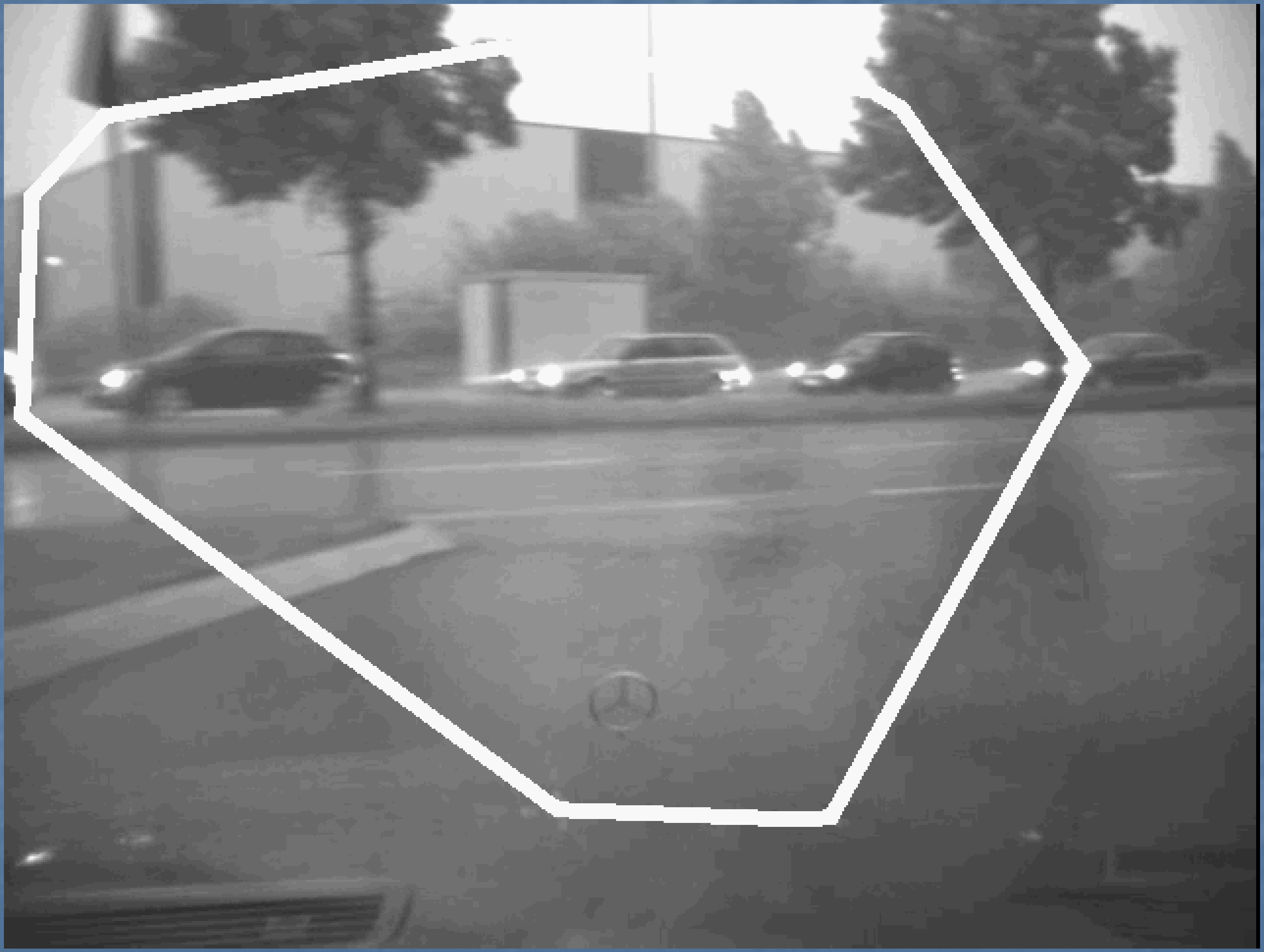
$$VC(N) = \frac{\left\| R \cap D_{N-T}^{N} \cap D_{N-T}^{N-2T} \right\|}{\left\| R \cap D_{N-T}^{N} \right\|}$$

Frame F_{N-T} with D_{N-T}^{N-2T}

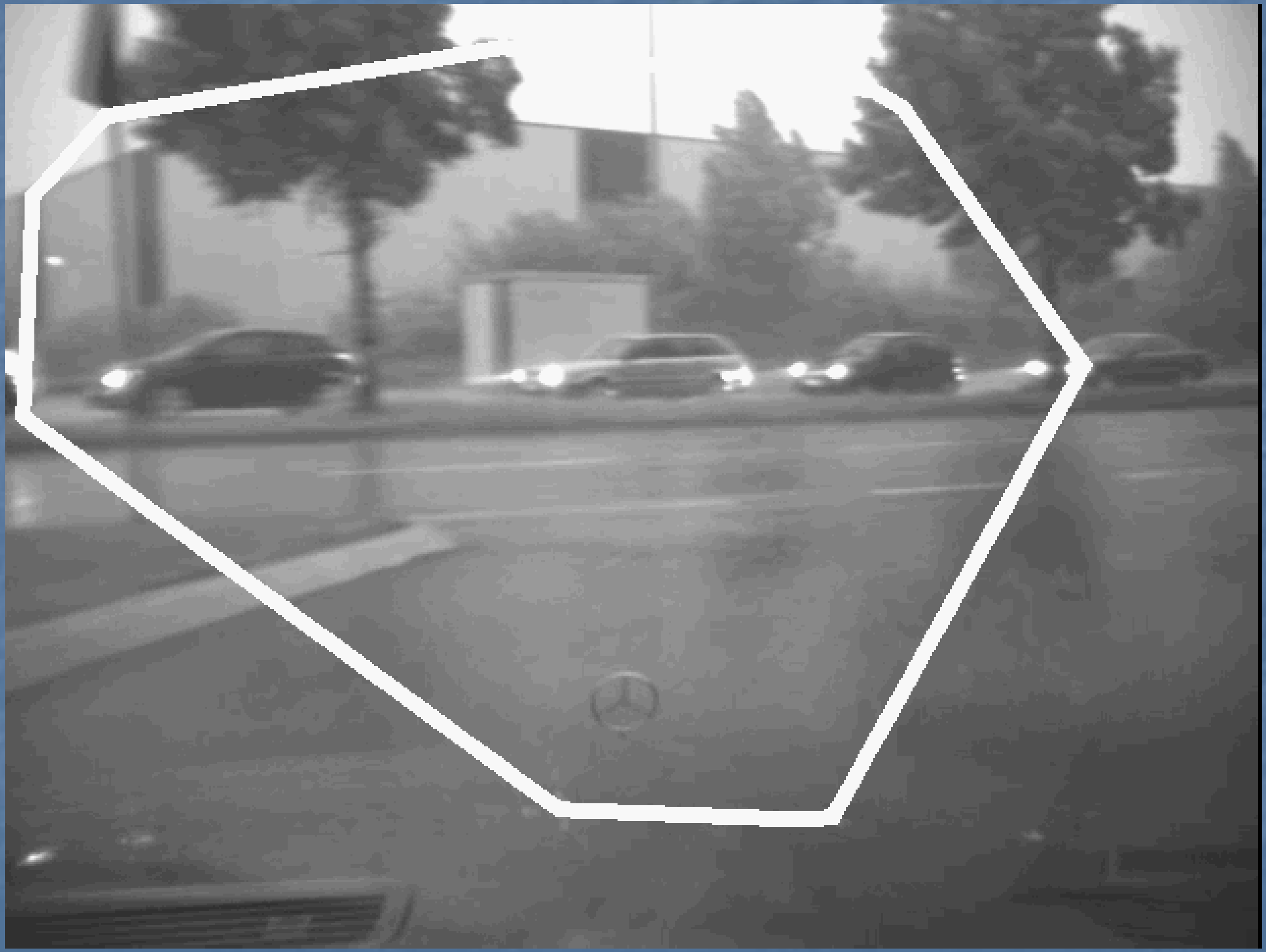
Frame F_{N-T} with D_{N-T}^{N}

Frame F_N with D_N^{N-T}

OBSERVATION: Visibility measures are inversely proportional to the speed!

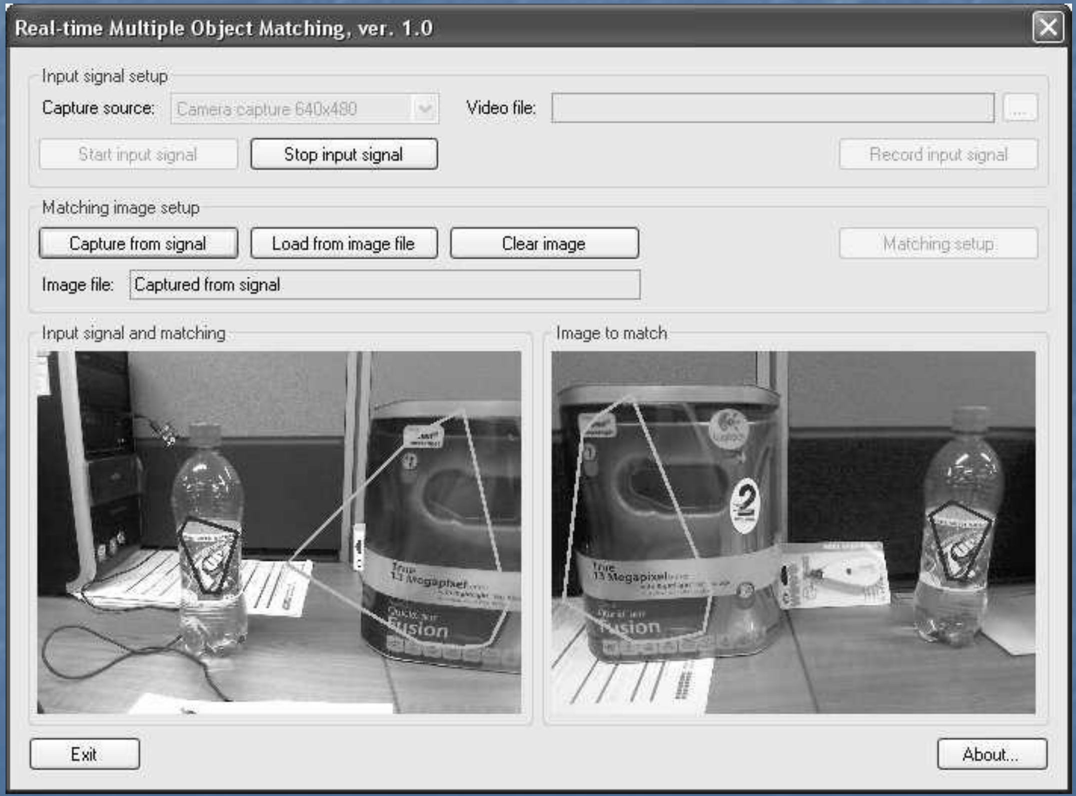


OBSERVATION: Visibility measures are inversely proportional to the speed!



Real-time implementation issues

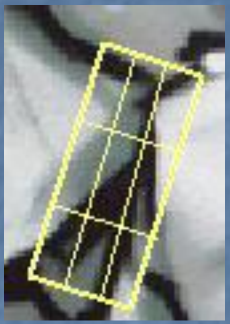
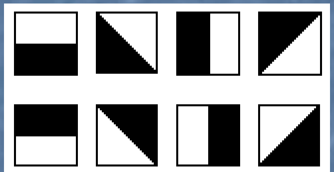
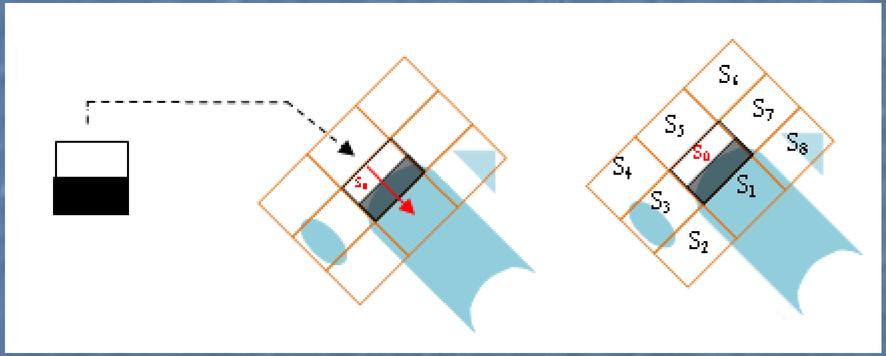
With certain simplifying assumptions, image-fragment matching can be run in real time (video-based) on a personal computer.



Supplementary techniques

An alternative keypoint descriptor of lower dimensionality (36)



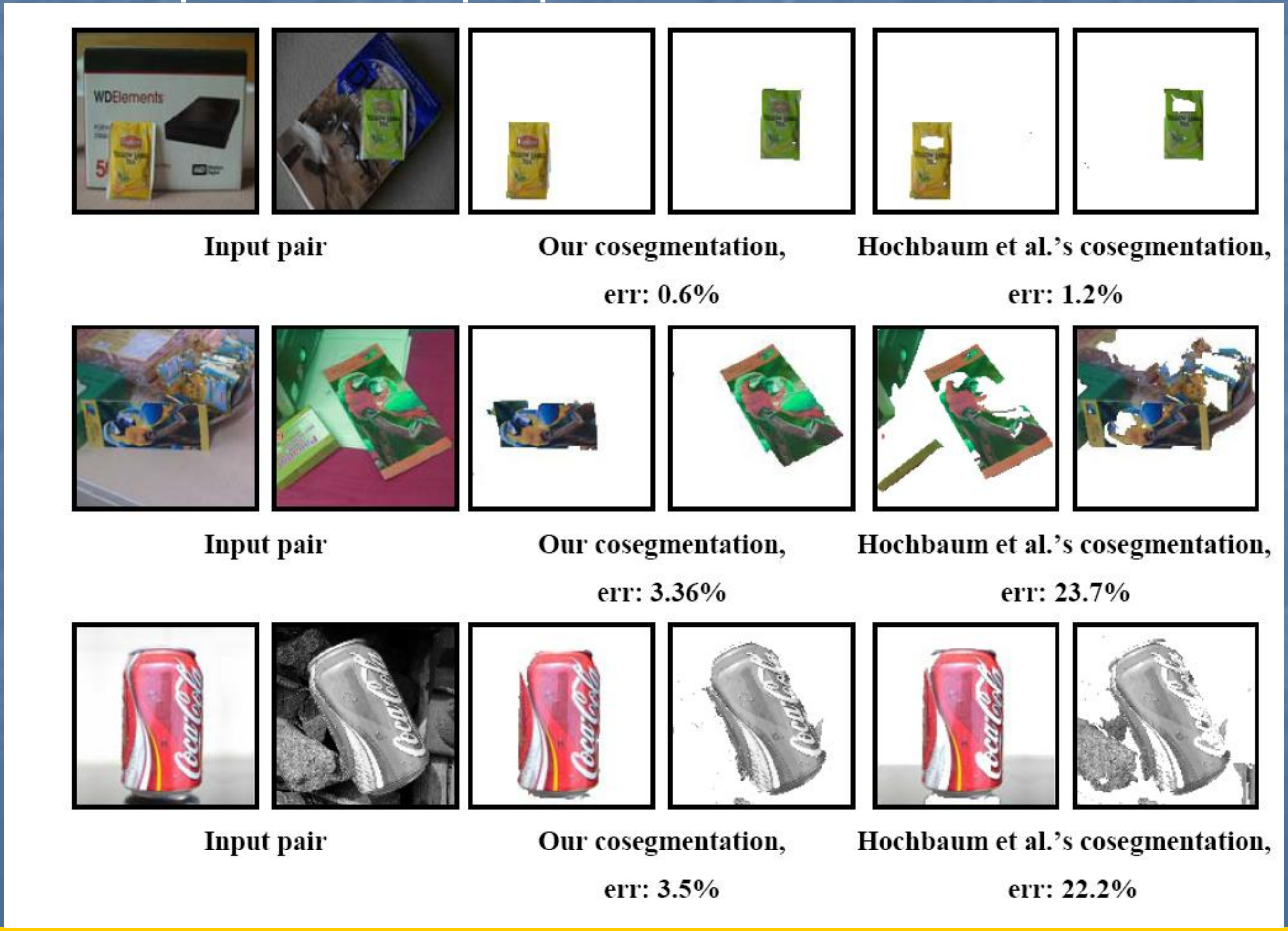
Almost as effective as SIFT for typically tested distortions.
Used in conjunction with TPS-based warping for matching nonlinearly distorted images/subimages.

A novel technique (Simplified Locality Sensitive Hashing - SLSH) implemented to reduce several times complexity of keypoint matching.

Accurate co-segmentation of near-duplicate fragments

Rectangular areas with near-duplicates detected can be used for shape refinements of these similar fragments.

A novel technique of co-segmentation (superior to published results) has been developed for this purpose.



Thank you !