ACCELERATING METRIC FILTERING BY IMPROVING BOUNDS ON ESTIMATED DISTANCES

VLADIMIR MIC AND PAVEL ZEZULA
FACULTY OF INFORMATICS, MASARYK UNIVERSITY, BRNO, CZECH REPUBLIC

MUNI C4E

sisap
Best Paper Award 2020

METRIC FILTERING

Topic

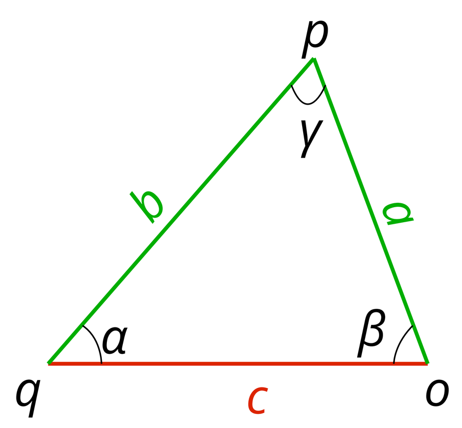
- Let q, o, p are objects in metric space (D, d)

  The upper-bound:
  $d(q,o) \leq d(q,p) + d(o,p)$

  The lower-bound:
  $d(q,o) \geq |d(q,p) - d(o,p)|$

Isometric Embedding to Euclidean Space

- The isometric embedding of q, o, p into 2D Euclidean space is always possible, → let's compute angles in the triangle
- Notation:



Tight Bounds on c?

- Two angles out of α, β, γ must be zero, and one 180° to have any of the bounds on c tight
- This practically never happens in contemporary data
- α, β, γ are always (nearly precisely) limited:

$$\Omega_{\min} \leq \alpha, \beta, \gamma \leq \Omega_{\max}$$

We enhance triangle inequalities with information about range of angles $\Omega_{\min}$, $\Omega_{\max}$ in triangles to improve the tightness of the bounds on unknown distances.

If the angles limitation $\Omega_{\min}, \Omega_{\max}$ holds, then newly derived bounds on c are correct.

NEWLY DERIVED BOUNDS ON C

New Bounds on c – Examples

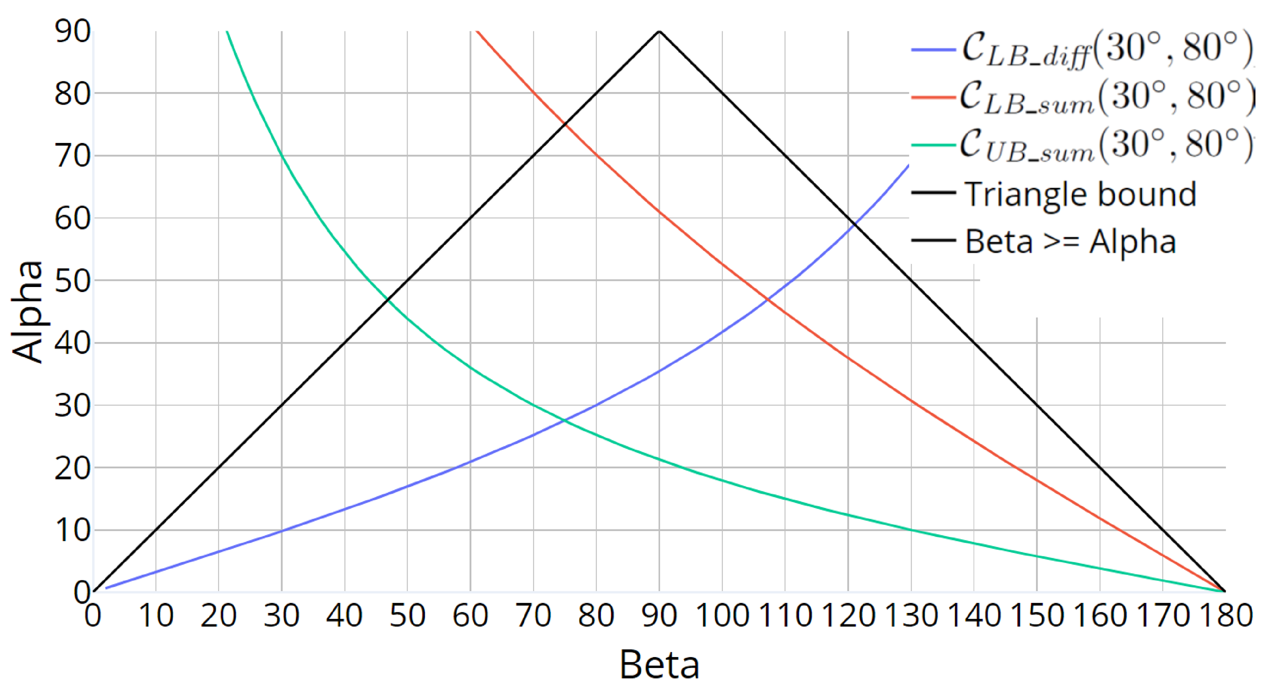
$[\Omega_{\min}, \Omega_{\max}]$	Upper-bound	Lower-bound	(Very new) lower-bound
$[0°, 180°]$	$c \leq (a+b) \cdot 1$	$c \geq \lvert a-b \rvert \cdot 1$	$c \geq (a+b) \cdot 0$
$[60°, 60°]$	$c \leq (a+b) \cdot 0.5$	*indef. exp.*	$c \geq (a+b) \cdot 0.5$
$[20°, 100°]$	$c \leq (a+b) \cdot 0.815$	$c \geq \lvert a-b \rvert \cdot 1.347$	$c \geq (a+b) \cdot 0.174$
$[20°, 80°]$	$c \leq (a+b) \cdot 0.742$	$c \geq \lvert a-b \rvert \cdot 1.532$	$c \geq (a+b) \cdot 0.174$
$[25°, 120°]$	$c \leq (a+b) \cdot 0.869$	$c \geq \lvert a-b \rvert \cdot 1.294$	$c \geq (a+b) \cdot 0.216$
$[25°, 90°]$	$c \leq (a+b) \cdot 0.752$	$c \geq \lvert a-b \rvert \cdot 1.570$	$c \geq (a+b) \cdot 0.216$
$[30°, 100°]$	$c \leq (a+b) \cdot 0.778$	$c \geq \lvert a-b \rvert \cdot 1.580$	$c \geq (a+b) \cdot 0.259$
$[30°, 80°]$	$c \leq (a+b) \cdot 0.684$	$c \geq \lvert a-b \rvert \cdot 1.938$	$c \geq (a+b) \cdot 0.259$

Consequences of Imprecise Angles Limitation $\Omega_{\min}, \Omega_{\max}$

- In practice, metric spaces guarantee just a trivial limitation $[\Omega_{\min}, \Omega_{\max}] = [0°, 180°]$
  - We still can set $[\Omega_{\min}, \Omega_{\max}]$ experimentally
  - Then, we lose the certainty that the bounds on c are correct

- We define the decomposition of all triangles into classes that correspond to the correctness of new bounds on distance c
- Each bound on c is either *'tight'*, *'correct but not tight'*, or *'wrong'*
- 15 classes of triangles exist (Some combinations do not exist)
- This allows to set $[\Omega_{\min}, \Omega_{\max}]$ to provide extremely precise bounds on c (more precisely: such that they almost do not lead to false negatives and false positives)

- Visualisation:



- x-axis expresses the size of angle β
- y-axis expresses the size of angle α
- We assume $\beta \geq \alpha$, and $\alpha + \beta \leq 180°$ without a loss of generality (see the video). Thus focus on a space below black lines
- Coloured curves assume $[\Omega_{\min}, \Omega_{\max}] = [30°, 80°]$
  - Decomposition of triangles according to the correctness of new bounds on c
  - All bounds on c are correct iff the point describing a triangle is between colour curves
- All triangles $\triangle a, b, c$ can be depicted as single points:

RESULTS

Visualisation of Real-life Data

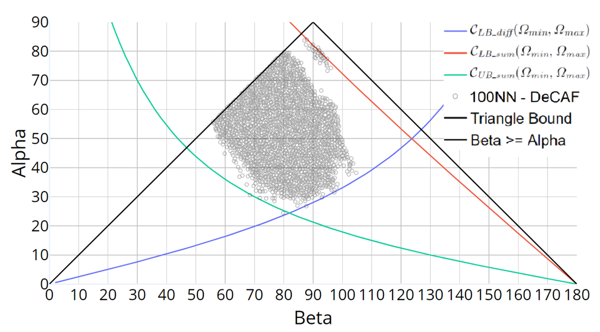


Fig.: DeCAF descriptors

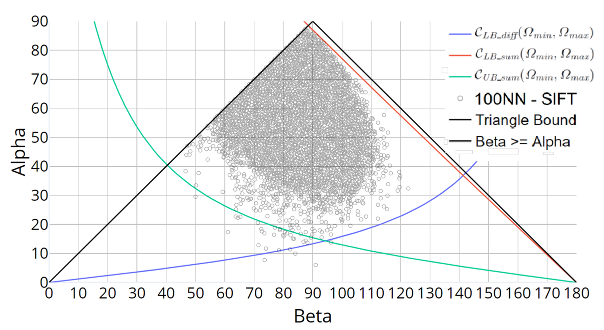


Fig.: SIFT descriptors

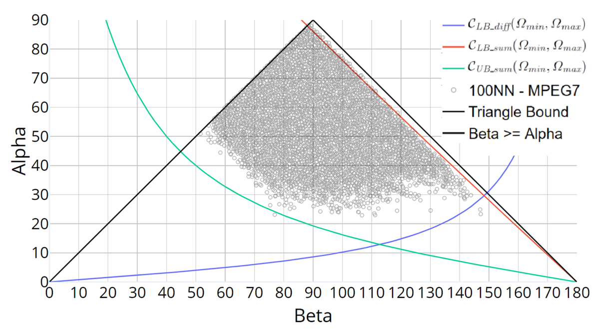


Fig.: MPEG7 descriptors

Black points: triangles sampled from the real life data, with focus on nearest neighbours.

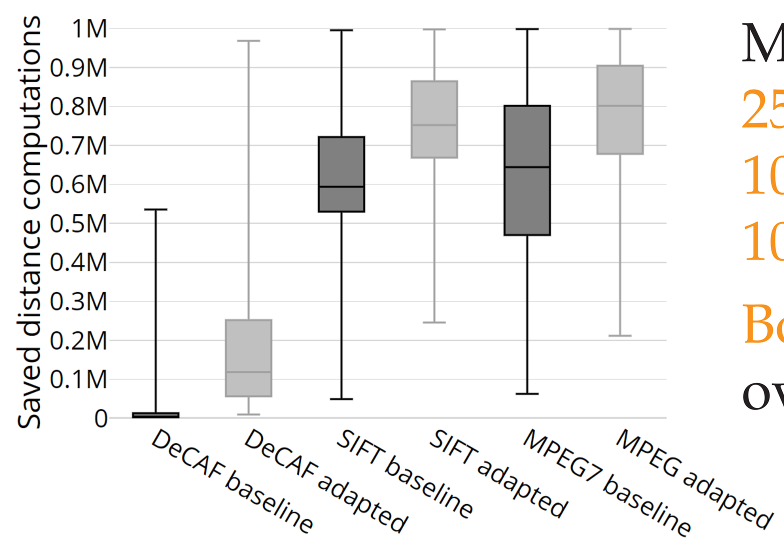
- Coloured curves visualise angles limitation $[\Omega_{\min}, \Omega_{\max}]$ used in experiments

Derived Bounds & Experiments

Selected angles limitation $[\Omega_{\min}, \Omega_{\max}]$, and new bounds on c

$[\Omega_{\min}, \Omega_{\max}]$	$[\Omega_{\min}, \Omega_{\max}]$	$[\Omega_{\min}, \Omega_{\max}]$
$[28°, 90°]$	$[8°, 86°]$	$[30°, 80°]$
$c \geq 1.66 \cdot \lvert b-a \rvert$	$c \geq 0.07 \cdot (b+a)$	$c \leq 0.68 \cdot (b+a)$
$[12°, 84°]$	$[3°, 88.5°]$	$[20°, 85°]$
$c \geq 1.26 \cdot \lvert b-a \rvert$	$c \geq 0.03 \cdot (b+a)$	$c \leq 0.76 \cdot (b+a)$
$[8°, 86°]$	$[4°, 88°]$	$[40°, 90°]$
$c \geq 1.16 \cdot \lvert b-a \rvert$	$c \geq 0.04 \cdot (b+a)$	$c \leq 0.71 \cdot (b+a)$

Increase of saved distance computations:



Metric filtering with 256 pivots, 1M data, 1000 query objects, 10NN queries.

Box-plots: values over particular query objects

Dark: pure triangle inequalities, light: new bounds

- The evaluation with new bounds is approximate. Nevertheless:
  - 999 out of 1000 answers to 10NN query contain ALL 10 NN in case of all datasets
  - The median query answer size is 10 in case of all datasets

https://disa.fi.muni.cz/~xmic/2020_09_Video_SISAP_Angles