MUNI C4E

Accelerating Metric Filtering by Improving Bounds on Estimated Distances

Vladimir Mic and Pavel Zezula
{xmic, zezula}@fi.muni.cz

Masaryk University, Brno, Czech Republic

September 30, 2020

Summary

- We enhance the definition of triangle inequalities to define tighter bounds on unknown distances
- Our approach is applicable for any metric space
- The metric filtering can be incorporated in a huge number of similarity indexes to speed-up the search practically for free
    - A straightforward example is given by the *PM-tree*

Summary

- We enhance the definition of triangle inequalities to define tighter bounds on unknown distances
- Our approach is applicable for any metric space
- The metric filtering can be incorporated in a huge number of similarity indexes to speed-up the search practically for free
  - A straightforward example is given by the *PM-tree*

Summary

- We enhance the definition of triangle inequalities to define tighter bounds on unknown distances
- Our approach is applicable for any metric space
- The metric filtering can be incorporated in a huge number of similarity indexes to speed-up the search practically for free
    - A straightforward example is given by the *PM-tree*

The Metric Space Similarity Model

- Domain of searched objects: D
- Similarity of two objects is expressed by a distance function d
  - $d : D \times D \mapsto \mathbf{R}_0^+$
  - The bigger the distance $d(x, y)$, the less similar objects x, y
- Similarity model: metric space (D, d)
  - $\forall x, y, z \in D$, the distance function d must satisfy:
  - $d(x, y) \geq 0$ (non-negativity)
  - $d(x, y) = d(y, x)$ (symmetry)
  - $d(x, y) = 0 \iff x = y$ (identity)
  - $d(x, y) + d(y, z) \geq d(x, z)$ (triangle inequality)

The Metric Space Similarity Model

- Domain of searched objects: D
- Similarity of two objects is expressed by a distance function d
  - $d : D \times D \mapsto \mathbf{R}_0^+$
  - The bigger the distance $d(x, y)$, the less similar objects x, y
- Similarity model: metric space (D, d)
  - $\forall x, y, z \in D$, the distance function d must satisfy:
  - $d(x, y) \geq 0$ (non-negativity)
  - $d(x, y) = d(y, x)$ (symmetry)
  - $d(x, y) = 0 \iff x = y$ (identity)
  - $d(x, y) + d(y, z) \geq d(x, z)$ (triangle inequality)

The Similarity Search

- $X \subseteq D$ is a searched dataset
- $q \in D$ is an arbitrary given query object
- The goal is to efficiently find the most similar objects to q: $Ans(q) \subseteq X$

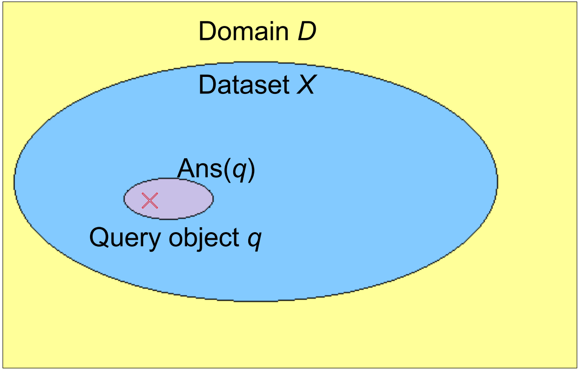


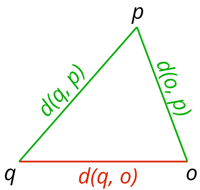
Fig.: Domain D, dataset X, query object q and answer set $Ans(q)$

Bounds on Unknown Distance

- Let $q, p, o \in D$ are arbitrary given objects
  - They are going to be the query object $q \in D$, object o from the dataset X, and a reference object $p \in D$, called *pivot*

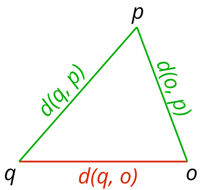
Bounds on Unknown Distance

- Let $q, p, o \in D$ are arbitrary given objects
  - They are going to be the query object $q \in D$, object o from the dataset X, and a reference object $p \in D$, called *pivot*
- Let us assume that distances $d(q, p)$ and $d(o, p)$ are known, and $d(q, o)$ is unknown



Bounds on Unknown Distance

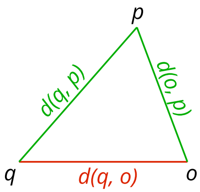
- Let $q, p, o \in D$ are arbitrary given objects
  - They are going to be the query object $q \in D$, object o from the dataset X, and a reference object $p \in D$, called *pivot*
- Let us assume that distances $d(q, p)$ and $d(o, p)$ are known, and $d(q, o)$ is unknown



- Triangle inequalities define the upper-bound and lower-bound on $d(q, o)$:

- $d(q, o) \leq d(q, p) + d(o, p)$
- $d(q, o) \geq |d(q, p) - d(o, p)|$

Metric Filtering



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

- Let us assume that we are interested just in objects within distance r

- If the upper-bound is smaller than r, o is guaranteed to be in the answer
- If the lower-bound is bigger than r, o cannot be in the answer

The Crucial Feature of the Bounds

The Key Question

How tight are these bounds in practice?

The Answer

The tightness of these bounds strongly suffers from the *dimensionality curse*

- the extreme case is the equilateral triangle: the lower-bound is 0, thus useless, the upper-bound is twice as big as $d(q, o)$

Our Contribution

- We enhance triangle inequalities with additional information to improve the tightness of the bounds on unknown distances
- Moreover, we define a new lower-bound on unknown distances

The Crucial Feature of the Bounds

The Key Question

How tight are these bounds in practice?

The Answer

The tightness of these bounds strongly suffers from the *dimensionality curse*

- the extreme case is the equilateral triangle: the lower-bound is 0, thus useless, the upper-bound is twice as big as $d(q, o)$

Our Contribution

- We enhance triangle inequalities with additional information to improve the tightness of the bounds on unknown distances
- Moreover, we define a new lower-bound on unknown distances

The Crucial Feature of the Bounds

The Key Question

How tight are these bounds in practice?

The Answer

The tightness of these bounds strongly suffers from the *dimensionality curse*

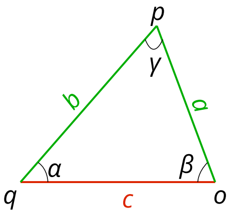
- the extreme case is the equilateral triangle: the lower-bound is 0, thus useless, the upper-bound is twice as big as $d(q, o)$

Our Contribution

- We enhance triangle inequalities with additional information to improve the tightness of the bounds on unknown distances
- Moreover, we define a new lower-bound on unknown distances

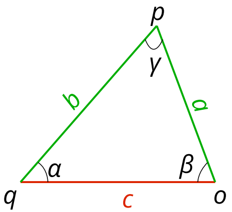
Embedding into the Euclidean Space

- The rule of triangle inequality allows to isometrically embed any three objects $q, o, p \in D$ into 2D Euclidean space
  - Here, we can use the cosine rule to evaluate angles in the triangle with sides of lengths $d(q, o), d(q, p), d(o, p)$
- We simplify the notation to have a triangle with sides $\triangle a, b, c$ and the corresponding angles α, β, γ
  - c is the unknown distance
  - We focus on a non-trivial case $a \neq 0, b \neq 0, c \neq 0$



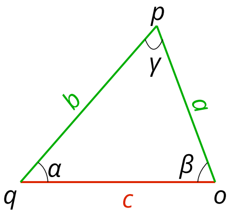
Embedding into the Euclidean Space

- The rule of triangle inequality allows to isometrically embed any three objects $q, o, p \in D$ into 2D Euclidean space
  - Here, we can use the cosine rule to evaluate angles in the triangle with sides of lengths $d(q, o)$, $d(q, p)$, $d(o, p)$
- We simplify the notation to have a triangle with sides $\triangle a, b, c$ and the corresponding angles α, β, γ
  - c is the unknown distance
  - We focus on a non-trivial case $a \neq 0, b \neq 0, c \neq 0$

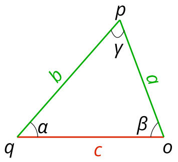


Embedding into the Euclidean Space

- The rule of triangle inequality allows to isometrically embed any three objects $q, o, p \in D$ into 2D Euclidean space
  - Here, we can use the cosine rule to evaluate angles in the triangle with sides of lengths $d(q, o)$, $d(q, p)$, $d(o, p)$
- We simplify the notation to have a triangle with sides $\triangle a, b, c$ and the corresponding angles α, β, γ
  - c is the unknown distance
  - We focus on a non-trivial case $a \neq 0, b \neq 0, c \neq 0$

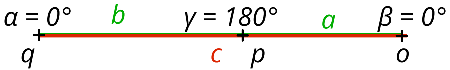


Tight Bounds?

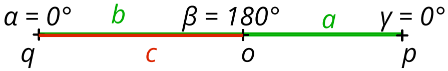


- When are the bounds on c tight?

- The tight upper-bound on c:
  $c = b + a$



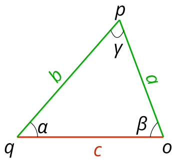
- The tight lower-bound on c:
  $c = |b - a|$



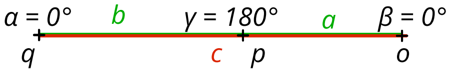
or a symmetric case with $\alpha = 180^\circ, \beta = \gamma = 0^\circ$

Tight Bounds?

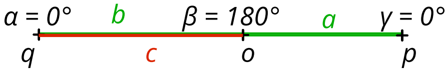
- When are the bounds on c tight?



- The tight upper-bound on c:
  $c = b + a$



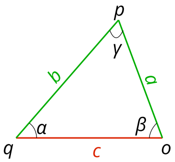
- The tight lower-bound on c:
  $c = |b - a|$



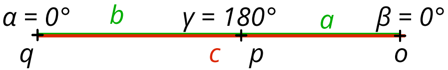
or a symmetric case with $\alpha = 180^\circ, \beta = \gamma = 0^\circ$

Tight Bounds?

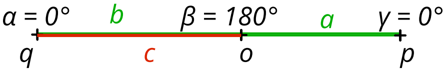
- When are the bounds on c tight?



- The tight upper-bound on c:
  $c = b + a$



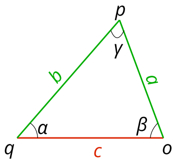
- The tight lower-bound on c:
  $c = |b - a|$



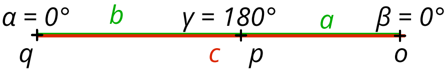
or a symmetric case with $\alpha = 180^\circ, \beta = \gamma = 0^\circ$

Tight Bounds?

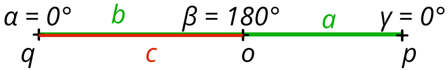
- When are the bounds on c tight?



- The tight upper-bound on c:
  $c = b + a$



- The tight lower-bound on c:
  $c = |b - a|$



or a symmetric case with $\alpha = 180^\circ, \beta = \gamma = 0^\circ$

The Answer

There must be two zero angles and the straight angle in a triangle $\triangle a, b, c$ to make any of these bounds on c tight

Tight Bounds?

There must be two zero angles and a straight angle in the triangle $\triangle a, b, c$ to make any of these bounds on c tight

The Question

How often this happens?

The Answer

Basically, never, in case of high dimensional data

Tight Bounds?

There must be two zero angles and a straight angle in the triangle $\triangle a, b, c$ to make any of these bounds on c tight

The Question

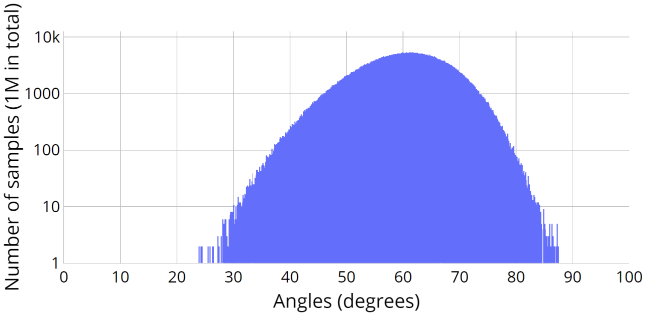
How often this happens?

The Answer

Basically, never, in case of high dimensional data

Example

- Example of angles α, β, γ in triangles $\triangle a, b, c$ (space of image visual descriptors[1])

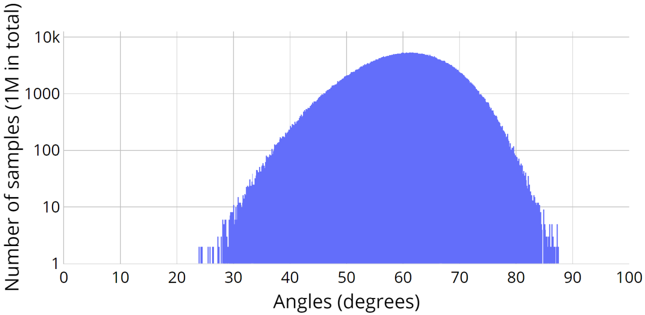


- Notice the log scale of y-axis (1 million angles in triangles $\triangle a, b, c$ are sampled)
- All angles α, β, γ are in range $[21^\circ, 91^\circ]$
- So, why we admit the whole range of angles $[0^\circ, 180^\circ]$ do define bounds on c?

[1]DeCAF descriptors, 4096 dimensional float vectors from a neural network, Euclidean space

Example

- Example of angles α, β, γ in triangles $\triangle a, b, c$ (space of image visual descriptors[1])



- Notice the log scale of y-axis (1 million angles in triangles $\triangle a, b, c$ are sampled)
- All angles α, β, γ are in range $[21^\circ, 91^\circ]$
- So, why we admit the whole range of angles $[0^\circ, 180^\circ]$ do define bounds on c?

[1] DeCAF descriptors, 4096 dimensional float vectors from a neural network, Euclidean space

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

…and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \qquad (1)$$

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

…and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos \gamma}{\cos \alpha + \cos \beta} \qquad (1)$$

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

- The lower-bound is:

$$c \geq |b - a|$$

...and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \qquad (1)$$

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \qquad (2)$$

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

- The lower-bound is:

$$c \geq |b - a|$$

…and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \qquad (1)$$

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \qquad (2)$$

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

- The lower-bound is:

$$c \geq |b - a|$$

...and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \quad (1)$$

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \quad (2)$$

Let us assume values $\Omega_{\text{min}}, \Omega_{\text{max}}$ such that for all triangles $\triangle a, b, c$ in the metric space:

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

- we prove that values $\Omega_{\text{min}}, \Omega_{\text{max}}$ limit the values of fractions in Eq. 1 and Eq. 2

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

- The lower-bound is:

$$c \geq |b - a|$$

…and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \qquad (1)$$

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \qquad (2)$$

Let us assume values $\Omega_{min}, \Omega_{max}$ such that for all triangles $\triangle a, b, c$ in the metric space:

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

- we prove that values $\Omega_{min}, \Omega_{max}$ limit the values of fractions in Eq. 1 and Eq. 2

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

- The lower-bound is:

$$c \geq |b - a|$$

…and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \quad (1)$$

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \quad (2)$$

Let us assume values $\Omega_{\text{min}}, \Omega_{\text{max}}$ such that for all triangles $\triangle a, b, c$ in the metric space:

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

- we prove that values $\Omega_{\text{min}}, \Omega_{\text{max}}$ limit the values of fractions in Eq. 1 and Eq. 2

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

- The lower-bound is:

$$c \geq |b - a|$$

…and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \qquad (1)$$

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \qquad (2)$$

Let us assume values $\Omega_{\text{min}}, \Omega_{\text{max}}$ such that for all triangles $\triangle a, b, c$ in the metric space:

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

- we prove that values $\Omega_{\text{min}}, \Omega_{\text{max}}$ limit the values of fractions in Eq. 1 and Eq. 2
- Notice that Ω_{min} and Ω_{max} always exist, as trivially $\Omega_{\text{min}} = 0^\circ$ and $\Omega_{\text{max}} = 180^\circ$
- Ω_{min} and Ω_{max} are meaningful if and only if $0^\circ \leq \Omega_{\text{min}} \leq 60^\circ \leq \Omega_{\text{max}} \leq 180^\circ$

Core of the Article

- The upper-bound is:

$$c \leq a + b$$

- The lower-bound is:

$$c \geq |b - a|$$

…and we prove equalities:

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \qquad (1)$$

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \qquad (2)$$

Let us assume values $\Omega_{\text{min}}, \Omega_{\text{max}}$ such that for all triangles $\triangle a, b, c$ in the metric space:

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

- we prove that values $\Omega_{\text{min}}, \Omega_{\text{max}}$ limit the values of fractions in Eq. 1 and Eq. 2
- Notice that Ω_{min} and Ω_{max} always exist, as trivially $\Omega_{\text{min}} = 0^\circ$ and $\Omega_{\text{max}} = 180^\circ$
- Ω_{min} and Ω_{max} are meaningful if and only if $0^\circ \leq \Omega_{\text{min}} \leq 60^\circ \leq \Omega_{\text{max}} \leq 180^\circ$

New Bounds on c

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \tag{1}$$

Let us denote:

- $\mathcal{C}_{\text{LB\_sum}}(\Omega_{\text{min}}, \Omega_{\text{max}})$ the minimum value of fraction in Equation 1
  - it defines the lower-bound on c based on the sum of a and b, since

$$c = (a + b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \geq (a + b) \cdot \mathcal{C}_{\text{LB\_sum}}(\Omega_{\text{min}}, \Omega_{\text{max}})$$

New Bounds on c

$$c = (a+b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \tag{1}$$

Let us denote:

- $\mathcal{C}_{\mathrm{LB\_sum}}(\Omega_{\mathrm{min}}, \Omega_{\mathrm{max}})$ the minimum value of fraction in Equation 1
  - it defines the lower-bound on c based on the sum of a and b, since

$$c = (a+b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \geq (a+b) \cdot \mathcal{C}_{\mathrm{LB\_sum}}(\Omega_{\mathrm{min}}, \Omega_{\mathrm{max}})$$

- $\mathcal{C}_{\mathrm{UB\_sum}}(\Omega_{\mathrm{min}}, \Omega_{\mathrm{max}})$ the maximum value of fraction in Equation 1
  - it defines the upper-bound on c based on the sum of a and b, since

$$c = (a+b) \cdot \frac{1 - \cos\gamma}{\cos\alpha + \cos\beta} \leq (a+b) \cdot \mathcal{C}_{\mathrm{UB\_sum}}(\Omega_{\mathrm{min}}, \Omega_{\mathrm{max}})$$

New Bounds on c

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \tag{2}$$

New Bounds on c

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{|\cos\alpha - \cos\beta|} \tag{2}$$

Let us denote:

- $\mathcal{C}_{\text{LB\_diff}}(\Omega_{\text{min}}, \Omega_{\text{max}})$ the minimum value of fraction in Equation 2
  - it defines the lower-bound on c based on the difference of a and b, since

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{\cos\alpha - \cos\beta} \geq |b - a| \cdot \mathcal{C}_{\text{LB\_diff}}(\Omega_{\text{min}}, \Omega_{\text{max}})$$

- $\mathcal{C}_{\text{UB\_diff}}(\Omega_{\text{min}}, \Omega_{\text{max}})$ the maximum value of fraction in Equation 2
  - it defines the upper-bound on c based on the difference of a and b, since

$$c = |b - a| \cdot \frac{1 + \cos\gamma}{\cos\alpha - \cos\beta} \leq |b - a| \cdot \mathcal{C}_{\text{UB\_diff}}(\Omega_{\text{min}}, \Omega_{\text{max}})$$

Four Bounds on c?

- We provide algorithms to evaluate all coefficients:
  - $\mathcal{C}_{LB\_sum}(\Omega_{min}, \Omega_{max})$
  - $\mathcal{C}_{UB\_sum}(\Omega_{min}, \Omega_{max})$
  - $\mathcal{C}_{LB\_diff}(\Omega_{min}, \Omega_{max})$
  - $\mathcal{C}_{UB\_diff}(\Omega_{min}, \Omega_{max})$

  ...for any range of permitted angles Ω_{min}, Ω_{max}

- Just $\mathcal{C}_{UB\_diff}(\Omega_{min}, \Omega_{max})$ is infinity for all meaningful ranges $[\Omega_{min}, \Omega_{max}]$
  - thus, $\mathcal{C}_{UB\_diff}(\Omega_{min}, \Omega_{max})$ provides just a trivial upper-bound infinity on c, for all meaningful ranges $[\Omega_{min}, \Omega_{max}]$
  - ...it is easy to prove that no other upper-bound on c can be defined using angles limitation $[\Omega_{min}, \Omega_{max}]$ and a difference of b and a. So we are correct.

New Bounds on c – Overview

Our Theoretical Contribution

We define two lower-bounds and an upper-bound on a distance c in a triangle $\triangle a, b, c$, that exploit

- the range of permitted angles $[\Omega_{min}, \Omega_{max}] : \Omega_{min} \leq \alpha, \beta, \gamma \leq \Omega_{max}$
- distances a and b *(their sum and diff, respectively)*

$[\Omega_{min}, \Omega_{max}]$	Upper-bound on c	Lower-bound on c	Lower-bound on c
$[0°, 180°]$	$c \leq (a+b) \cdot 1$	$c \geq \lvert a-b \rvert \cdot 1$	$c \geq (a+b) \cdot 0$
$[60°, 60°]$	$c \leq (a+b) \cdot 0.5$	*undefined*[2]	$c \geq (a+b) \cdot 0.5$

Table: Trivial examples of bounds on c for given ranges of angles $\Omega_{min}, \Omega_{max}$

[2] indefinite expression $0 \cdot \infty$

New Bounds on c – Overview

Our Theoretical Contribution

We define two lower-bounds and an upper-bound on a distance c in a triangle $\triangle a, b, c$, that exploit

- the range of permitted angles $[\Omega_{min}, \Omega_{max}] : \Omega_{min} \leq \alpha, \beta, \gamma \leq \Omega_{max}$
- distances a and b *(their sum and diff, respectively)*

$[\Omega_{min}, \Omega_{max}]$	Upper-bound on c	Lower-bound on c	Lower-bound on c
$[0^\circ, 180^\circ]$	$c \leq (a+b) \cdot 1$	$c \geq \lvert a-b \rvert \cdot 1$	$c \geq (a+b) \cdot 0$
$[60^\circ, 60^\circ]$	$c \leq (a+b) \cdot 0.5$	*undefined*[2]	$c \geq (a+b) \cdot 0.5$

Table: Trivial examples of bounds on c for given ranges of angles $\Omega_{min}, \Omega_{max}$

[2] indefinite expression $0 \cdot \infty$

New Bounds on c – Overview

Our Theoretical Contribution

We define two lower-bounds and an upper-bound on a distance c in a triangle $\triangle a, b, c$, that exploit

- the range of permitted angles $[\Omega_{min}, \Omega_{max}] : \Omega_{min} \leq \alpha, \beta, \gamma \leq \Omega_{max}$
- distances a and b *(their sum and diff, respectively)*

$[\Omega_{min}, \Omega_{max}]$	Upper-bound on c	Lower-bound on c	Lower-bound on c
$[0°, 180°]$	$c \leq (a + b) \cdot 1$	$c \geq \lvert a - b \rvert \cdot 1$	$c \geq (a + b) \cdot 0$
$[60°, 60°]$	$c \leq (a + b) \cdot 0.5$	*undefined*[2]	$c \geq (a + b) \cdot 0.5$

Table: Trivial examples of bounds on c for given ranges of angles Ω_{min}, Ω_{max}

[2] indefinite expression $0 \cdot \infty$

New Bounds on c – Overview

Our Theoretical Contribution

We define two lower-bounds and an upper-bound on a distance c in a triangle $\triangle a, b, c$, that exploit

- the range of permitted angles $[\Omega_{min}, \Omega_{max}] : \Omega_{min} \leq \alpha, \beta, \gamma \leq \Omega_{max}$
- distances a and b *(their sum and diff, respectively)*

$[\Omega_{min}, \Omega_{max}]$	Upper-bound on c	Lower-bound on c	Lower-bound on c
$[0°, 180°]$	$c \leq (a + b) \cdot 1$	$c \geq \lvert a - b \rvert \cdot 1$	$c \geq (a + b) \cdot 0$
$[60°, 60°]$	$c \leq (a + b) \cdot 0.5$	*undefined*[2]	$c \geq (a + b) \cdot 0.5$

Table: Trivial examples of bounds on c for given ranges of angles $\Omega_{min}, \Omega_{max}$

[2] indefinite expression $0 \cdot \infty$

New Bounds on c – Examples

- The original upper-bound: $c \leq a + b$
- The original lower-bound: $c \geq |a - b|$

$[\Omega_{min}, \Omega_{max}]$	Upper-bound on c	Lower-bound on c	Lower-bound on c
$[20°, 100°]$	$c \leq (a+b) \cdot 0.815$	$c \geq \|a-b\| \cdot 1.347$	$c \geq (a+b) \cdot 0.174$
$[20°, 80°]$	$c \leq (a+b) \cdot 0.742$	$c \geq \|a-b\| \cdot 1.532$	$c \geq (a+b) \cdot 0.174$
$[25°, 120°]$	$c \leq (a+b) \cdot 0.869$	$c \geq \|a-b\| \cdot 1.294$	$c \geq (a+b) \cdot 0.216$
$[25°, 90°]$	$c \leq (a+b) \cdot 0.752$	$c \geq \|a-b\| \cdot 1.570$	$c \geq (a+b) \cdot 0.216$
$[30°, 100°]$	$c \leq (a+b) \cdot 0.778$	$c \geq \|a-b\| \cdot 1.580$	$c \geq (a+b) \cdot 0.259$
$[30°, 80°]$	$c \leq (a+b) \cdot 0.684$	$c \geq \|a-b\| \cdot 1.938$	$c \geq (a+b) \cdot 0.259$

Table: Non-trivial examples of bounds on c for given ranges of angles Ω_{min}, Ω_{max}

New Bounds on c – Examples

- The original upper-bound: $c \leq a + b$
- The original lower-bound: $c \geq |a - b|$

$[\Omega_{min}, \Omega_{max}]$	Upper-bound on c	Lower-bound on c	Lower-bound on c
$[20°, 100°]$	$c \leq (a+b) \cdot 0.815$	$c \geq \|a-b\| \cdot 1.347$	$c \geq (a+b) \cdot 0.174$
$[20°, 80°]$	$c \leq (a+b) \cdot 0.742$	$c \geq \|a-b\| \cdot 1.532$	$c \geq (a+b) \cdot 0.174$
$[25°, 120°]$	$c \leq (a+b) \cdot 0.869$	$c \geq \|a-b\| \cdot 1.294$	$c \geq (a+b) \cdot 0.216$
$[25°, 90°]$	$c \leq (a+b) \cdot 0.752$	$c \geq \|a-b\| \cdot 1.570$	$c \geq (a+b) \cdot 0.216$
$[30°, 100°]$	$c \leq (a+b) \cdot 0.778$	$c \geq \|a-b\| \cdot 1.580$	$c \geq (a+b) \cdot 0.259$
$[30°, 80°]$	$c \leq (a+b) \cdot 0.684$	$c \geq \|a-b\| \cdot 1.938$	$c \geq (a+b) \cdot 0.259$

Table: Non-trivial examples of bounds on c for given ranges of angles $\Omega_{min}, \Omega_{max}$

How to Set Angles Limitation

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

How to Set Angles Limitation

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

How to Set Angles Limitation

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

- On the consequences of wrong angles limitation $[\Omega_{min}, \Omega_{max}]$:
  1. We prove and discuss that each of new three bounds on c can be correct even in a triangle $\triangle a, b, c$ where angles α, β, γ violate the angles limitation $[\Omega_{min}, \Omega_{max}]$
  2. Moreover, we prove that all three bounds on c can be correct at the same time in a triangle $\triangle a, b, c$ where angles α, β, γ violate the angles limitation $[\Omega_{min}, \Omega_{max}]$

How to Set Angles Limitation

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

- On the consequences of wrong angles limitation $[\Omega_{\text{min}}, \Omega_{\text{max}}]$:
  1. We prove and discuss that each of new three bounds on c can be correct even in a triangle $\triangle a, b, c$ where angles α, β, γ violate the angles limitation $[\Omega_{\text{min}}, \Omega_{\text{max}}]$
  2. Moreover, we prove that all three bounds on c can be correct at the same time in a triangle $\triangle a, b, c$ where angles α, β, γ violate the angles limitation $[\Omega_{\text{min}}, \Omega_{\text{max}}]$

When are the Bounds on c Correct?

- So, we derive equations – one for each of three new bounds on c – to define
  - triangles for which a bound on c is wrong,
  - correct, and
  - tight,

  respectively. (We provide them just in the article.)

[3] Angle γ is not needed since it is given as $\gamma = 180^\circ - \alpha - \beta$

When are the Bounds on *c* Correct?

- So, we derive equations – one for each of three new bounds on *c* – to define
  - triangles for which a bound on *c* is wrong,
  - correct, and
  - tight,

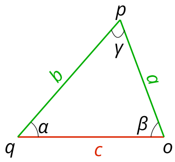
  respectively. (We provide them just in the article.)
- The correctness of the new bounds on *c* is given just by its angles α and β[3]
  - This enables us to clearly visualise the classes of triangles

[3] Angle γ is not needed since it is given as $\gamma = 180^\circ - \alpha - \beta$

When are the Bounds on c Correct?

- So, we derive equations – one for each of three new bounds on c – to define
  - triangles for which a bound on c is wrong,
  - correct, and
  - tight,

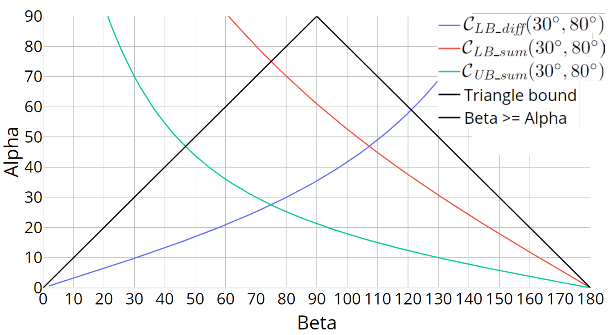
  respectively. (We provide them just in the article.)
- The correctness of the new bounds on c is given just by its angles α and β[3]
  - This enables us to clearly visualise the classes of triangles
- From now on, we assume $b \geq a$ in a triangle $\triangle a, b, c$, without a loss of generality for the similarity search, and for the sake of simplicity:
  - If $b < a$, then swapping of the notation of objects q and o swaps the distances b and a, and preserves the distance c thanks to the symmetry $d(q, o) = d(o, q)$. Thus, we can assume $b \geq a$ as far as we focus on an isolated triangle.



[3] Angle γ is not needed since it is given as $\gamma = 180° - \alpha - \beta$

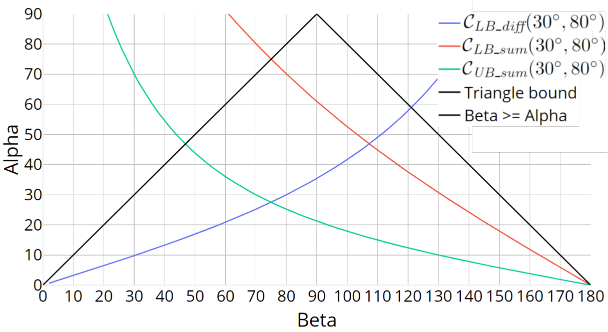
When are the Bounds on *c* Correct?

- We introduce plots like this one to depict all possible triangles $\triangle a, b, c$ as single points:
  - *x*-axis expresses the size of angle β
  - *y*-axis expresses the size of angle α
  - Since $\beta \geq \alpha$, and $\alpha + \beta \leq 180°$, we are interested just in points bellow the solid black lines *(…so please focus on the triangular shape in the plot)*

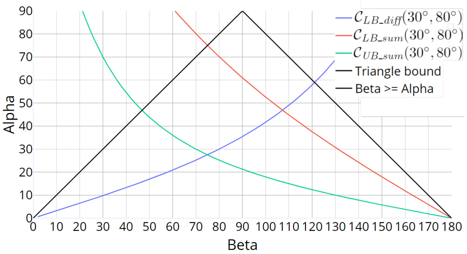


When are the Bounds on c Correct?

- We introduce plots like this one to depict all possible triangles $\triangle a, b, c$ as single points:
  - x-axis expresses the size of angle β
  - y-axis expresses the size of angle α
  - Since $\beta \geq \alpha$, and $\alpha + \beta \leq 180°$, we are interested just in points bellow the solid black lines *(...so please focus on the triangular shape in the plot)*

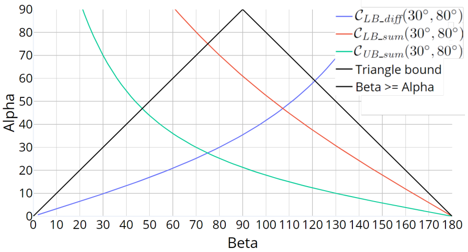


When are the Bounds on c Correct?



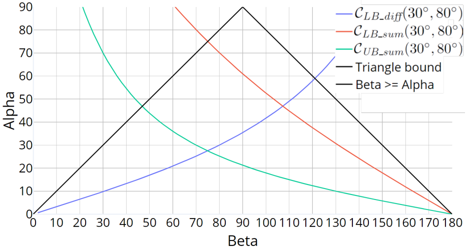
- Let $[\Omega_{\text{min}}, \Omega_{\text{max}}]$ is a given. For example: $[\Omega_{\text{min}}, \Omega_{\text{max}}] = [30°, 80°]$
- We define blue, green, and orange functions, which are depicted for $[\Omega_{\text{min}}, \Omega_{\text{max}}] = [30°, 80°]$ in the figure
  - If a triangle $\triangle a, b, c$ is represented by a point above or on a blue curve, then the lower-bound $c \geq |b - a| \cdot \mathcal{C}_{\text{LB\_diff}}(\Omega_{\text{min}}, \Omega_{\text{max}})$ is correct
  - If the triangle is depicted below a blue curve, then this lower-bound is wrong
  - Iff the triangle is depicted on a blue curve, then this lower-bound is tight

When are the Bounds on c Correct?



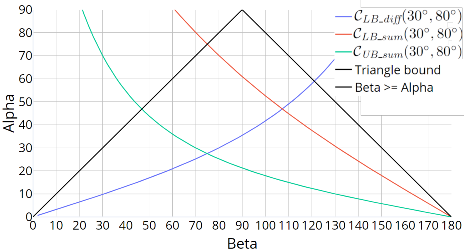
- Example for $[\Omega_{\text{min}}, \Omega_{\text{max}}] = [30°, 80°]$
  - If a triangle $\triangle a, b, c$ is represented by a point above or on a green curve, then the upper-bound $c \leq (a + b) \cdot \mathcal{C}_{\text{UB\_sum}}(\Omega_{\text{min}}, \Omega_{\text{max}})$ is correct
  - If the triangle is depicted below a green curve, then this upper-bound is wrong
  - Iff the triangle is depicted on a green curve, then this upper-bound is tight

When are the Bounds on c Correct?



- Example for $[\Omega_{\text{min}}, \Omega_{\text{max}}] = [30°, 80°]$
  - If a triangle $\triangle a, b, c$ is represented by a point *below*(!) or on an orange curve, then the lower-bound $c \geq (a + b) \cdot \mathcal{C}_{\text{LB\_sum}}(\Omega_{\text{min}}, \Omega_{\text{max}})$ is correct
  - If the triangle is depicted above a orange curve, then this lower-bound is wrong
  - Iff the triangle is depicted on a orange curve, then this lower-bound is tight

Conclusion on Setting Angles Limitation



- We need to set angles limitation $[\Omega_{\min}, \Omega_{\max}]$ to tightly embrace the points representing triangles $\triangle a, b, c$ drawn from the metric space by blue, green, and orange curves in a figure made for $[\Omega_{\min}, \Omega_{\max}]$[4]

[4] formal definition is in the article

Visualisations of Real Life Data

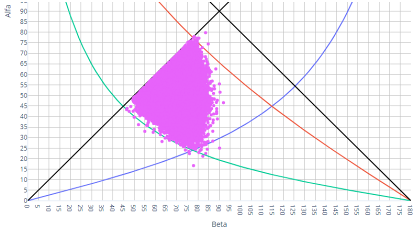


Fig.: DeCAF, $[\Omega_{min}, \Omega_{max}] = [25°, 80°]$

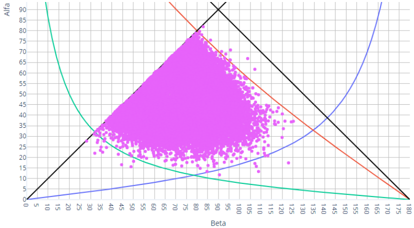


Fig.: MPEG7, $[\Omega_{min}, \Omega_{max}] = [20°, 110°]$

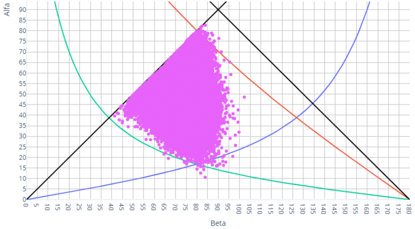


Fig.: SIFT, $[\Omega_{min}, \Omega_{max}] = [20°, 90°]$

- Purple points represent randomly selected triangles $\triangle a, b, c$
- It is possible to set $[\Omega_{min}, \Omega_{max}]$ to tightly embrace depicted points
- However, we are interested in the similarity search, i.e. in triangles with extremely small distance c

Visualisations of Real Life Data

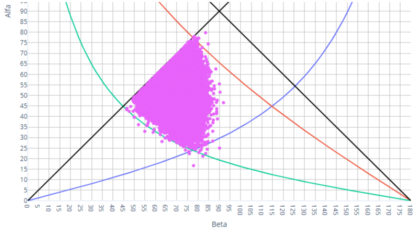


Fig.: DeCAF, $[\Omega_{min}, \Omega_{max}] = [25°, 80°]$

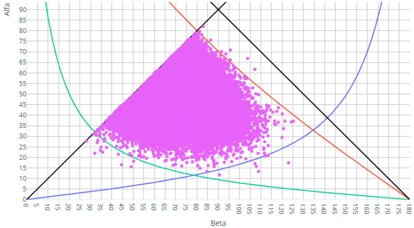


Fig.: MPEG7, $[\Omega_{min}, \Omega_{max}] = [20°, 110°]$

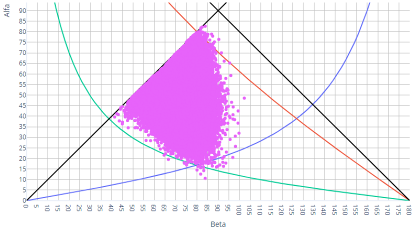
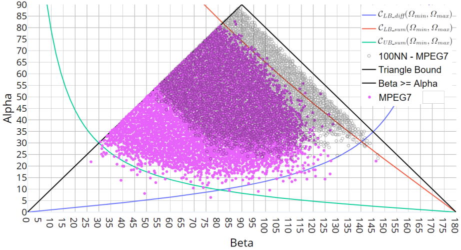


Fig.: SIFT, $[\Omega_{min}, \Omega_{max}] = [20°, 90°]$

- Purple points represent randomly selected triangles $\triangle a, b, c$
- It is possible to set $[\Omega_{min}, \Omega_{max}]$ to tightly embrace depicted points
- However, we are interested in the similarity search, i.e. in triangles with extremely small distance c

Triangles with Small Distance c

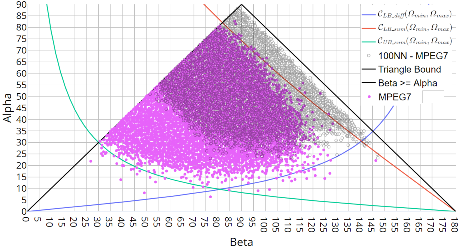
- Triangles $\triangle a, b, c$ with extremely small c form a different distribution:



- Purple points: randomly selected triangles
- Black circles: triangles where $c = d(q, o)$ is the distance of a random $q \in D$ to each one of its 100 nearest neighbours $o \in X$, for 1000 different $q \in D$

Triangles with Small Distance c

- Triangles $\triangle a, b, c$ with extremely small c form a different distribution:



- Purple points: randomly selected triangles
- Black circles: triangles where $c = d(q, o)$ is the distance of a random $q \in D$ to each one of its 100 nearest neighbours $o \in X$, for 1000 different $q \in D$
  - Angles limitation $[\Omega_{\min}, \Omega_{\max}]$ must be selected independently for each new bound on c to tightly embrace the black points depicting triangles with a near neighbour
  - Future work: We would really like to know, how to set $[\Omega_{\min}, \Omega_{\max}]$ for each bound on c analytically, without using these plots

Visualisations of Real Life Data

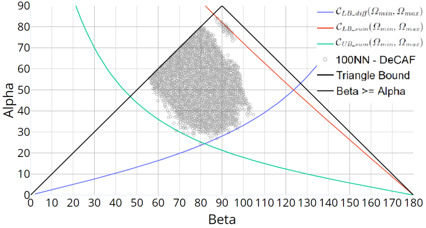


Fig.: DeCAF descriptors

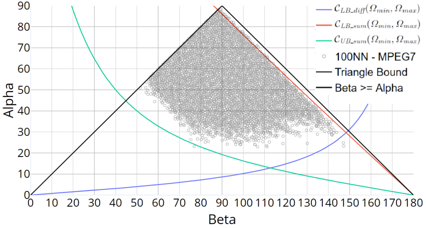


Fig.: MPEG7 descriptors

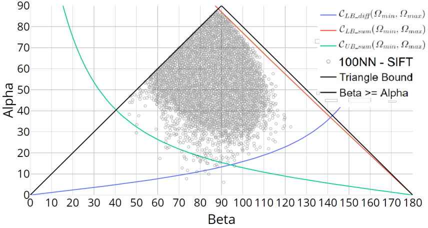


Fig.: SIFT descriptors

Derived Bounds on c

	$\mathcal{C}_{LB\_diff}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{LB\_sum}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{UB\_sum}(\Omega_{min}, \Omega_{max})$
DeCAF descriptors	$[28°, 90°]$	$[8°, 86°]$	$[30°, 80°]$
	$c \geq 1.664 \cdot \lvert b - a \rvert$	$c \geq 0.070 \cdot (b + a)$	$c \leq 0.684 \cdot (b + a)$
SIFT descriptors	$[12°, 84°]$	$[3°, 88.5°]$	$[20°, 85°]$
	$c \geq 1.264 \cdot \lvert b - a \rvert$	$c \geq 0.026 \cdot (b + a)$	$c \leq 0.762 \cdot (b + a)$
MPEG7 descriptors	$[8°, 86°]$	$[4°, 88°]$	$[40°, 90°]$
	$c \geq 1.162 \cdot \lvert b - a \rvert$	$c \geq 0.035 \cdot (b + a)$	$c \leq 0.710 \cdot (b + a)$

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

- So how the new bounds on c infer the metric filtering in practice?

Derived Bounds on c

	$\mathcal{C}_{LB\_diff}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{LB\_sum}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{UB\_sum}(\Omega_{min}, \Omega_{max})$
DeCAF descriptors	$[28°, 90°]$	$[8°, 86°]$	$[30°, 80°]$
	$c \geq 1.664 \cdot \|b - a\|$	$c \geq 0.070 \cdot (b + a)$	$c \leq 0.684 \cdot (b + a)$
SIFT descriptors	$[12°, 84°]$	$[3°, 88.5°]$	$[20°, 85°]$
	$c \geq 1.264 \cdot \|b - a\|$	$c \geq 0.026 \cdot (b + a)$	$c \leq 0.762 \cdot (b + a)$
MPEG7 descriptors	$[8°, 86°]$	$[4°, 88°]$	$[40°, 90°]$
	$c \geq 1.162 \cdot \|b - a\|$	$c \geq 0.035 \cdot (b + a)$	$c \leq 0.710 \cdot (b + a)$

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

- So how the new bounds on c infer the metric filtering in practice?

Derived Bounds on c

	$\mathcal{C}_{LB\_diff}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{LB\_sum}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{UB\_sum}(\Omega_{min}, \Omega_{max})$
DeCAF descriptors	$[28°, 90°]$	$[8°, 86°]$	$[30°, 80°]$
	$c \geq 1.664 \cdot \lvert b - a \rvert$	$c \geq 0.070 \cdot (b + a)$	$c \leq 0.684 \cdot (b + a)$
SIFT descriptors	$[12°, 84°]$	$[3°, 88.5°]$	$[20°, 85°]$
	$c \geq 1.264 \cdot \lvert b - a \rvert$	$c \geq 0.026 \cdot (b + a)$	$c \leq 0.762 \cdot (b + a)$
MPEG7 descriptors	$[8°, 86°]$	$[4°, 88°]$	$[40°, 90°]$
	$c \geq 1.162 \cdot \lvert b - a \rvert$	$c \geq 0.035 \cdot (b + a)$	$c \leq 0.710 \cdot (b + a)$

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

- So how the new bounds on c infer the metric filtering in practice?

Derived Bounds on c

	$\mathcal{C}_{LB\_diff}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{LB\_sum}(\Omega_{min}, \Omega_{max})$	$\mathcal{C}_{UB\_sum}(\Omega_{min}, \Omega_{max})$
DeCAF descriptors	$[28°, 90°]$	$[8°, 86°]$	$[30°, 80°]$
	$c \geq 1.664 \cdot \|b - a\|$	$c \geq 0.070 \cdot (b + a)$	$c \leq 0.684 \cdot (b + a)$
SIFT descriptors	$[12°, 84°]$	$[3°, 88.5°]$	$[20°, 85°]$
	$c \geq 1.264 \cdot \|b - a\|$	$c \geq 0.026 \cdot (b + a)$	$c \leq 0.762 \cdot (b + a)$
MPEG7 descriptors	$[8°, 86°]$	$[4°, 88°]$	$[40°, 90°]$
	$c \geq 1.162 \cdot \|b - a\|$	$c \geq 0.035 \cdot (b + a)$	$c \leq 0.710 \cdot (b + a)$

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

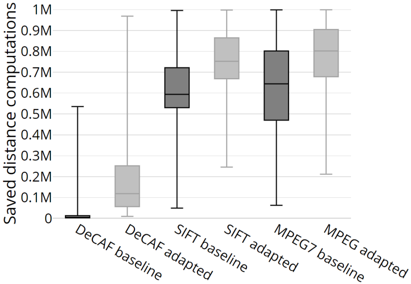
- So how the new bounds on c infer the metric filtering in practice?

Experiments

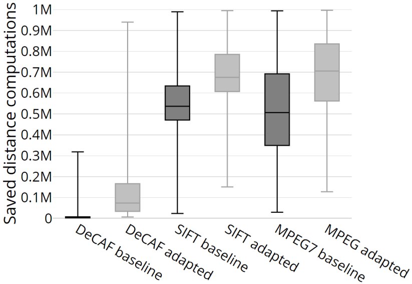
- We examine number of saved distance computations with normal triangle inequalities and with new bounds
  - Filtering with 256 random pivots
  - Evaluation of 1000 random query objects – values depicted by box-plots
  - Searching for $k \in 10, 100$ nearest neighbours (*NN*)
  - *k*-NN queries are evaluated with gradual shrinking of the searching radius
  - 3 datasets of size 1M objects each: SIFTs, DeCAF, MPEG7

Experiments

- Dark and light box-plots describe filtering with pure triangle inequalities, and with new bounds, respectively
- Median numbers of skipped distance computations in case of 10NN queries increase:
  - from 0.4 % to 11.8 % (DeCAF)
  - from 59.4 % to 75.2 % (SIFT)
  - from 64.5 % to 80.2 % (MPEG7)



(a) 10NN



(b) 100NN

Quality of the Approximation

- The evaluation with new bounds is just approximate
  - Answers can contain just some of the true nearest neighbours,
  - and also false positives due to a wrong upper-bound or gradual shrinking of the searching radius.

- 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

  - Answers to 100NN queries contain all 100 NN in case of 955, 963 and 923 query objects in case of the DeCAF, SIFT, and MPEG7 dataset, respectively.
  - The biggest answers are in case of MPEG7 descriptors: 105 objects on median

Quality of the Approximation

- The evaluation with new bounds is just approximate
  - Answers can contain just some of the true nearest neighbours,
  - and also false positives due to a wrong upper-bound or gradual shrinking of the searching radius.

- 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

  - Answers to 100NN queries contain all 100 NN in case of 955, 963 and 923 query objects in case of the DeCAF, SIFT, and MPEG7 dataset, respectively.
  - The biggest answers are in case of MPEG7 descriptors: 105 objects on median

Quality of the Approximation

- The evaluation with new bounds is just approximate
  - Answers can contain just some of the true nearest neighbours,
  - and also false positives due to a wrong upper-bound or gradual shrinking of the searching radius.

- 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

  - Answers to 100NN queries contain all 100 NN in case of 955, 963 and 923 query objects in case of the DeCAF, SIFT, and MPEG7 dataset, respectively.
  - The biggest answers are in case of MPEG7 descriptors: 105 objects on median

MUNI
C4E

EUROPEAN UNION
European Structural and Investment Funds
Operational Programme Research,
Development and Education

MINISTRY OF EDUCATION,
YOUTH AND SPORTS