

Searching similar images in PostgreSQL

Alexander Korotkov
NRNU MEPhI

It works!

Query:



Similar:



It works!

Query:



Similar:



It works!

Query:



Similar:



It works!

Query:



Similar:



It works!

Query:



Similar:



It works!

Query:



Similar:



It works!

Query:



Similar:



Sometimes not so perfect...

Query:



Similar:



How it works?

Process images and create an index.

```
CREATE TABLE pat AS (  
    SELECT  
        id,  
        shuffle_pattern(pattern) AS pattern,  
        pattern2signature(pattern) AS signature  
    FROM (  
        SELECT  
            id,  
            jpeg2pattern(data) AS pattern  
        FROM  
            image  
    ) x  
);  
CREATE INDEX pat_signature_idx ON pat USING gist (signature);  
CREATE INDEX pat_id_idx ON pat(id);
```

How it works?

Search similar images

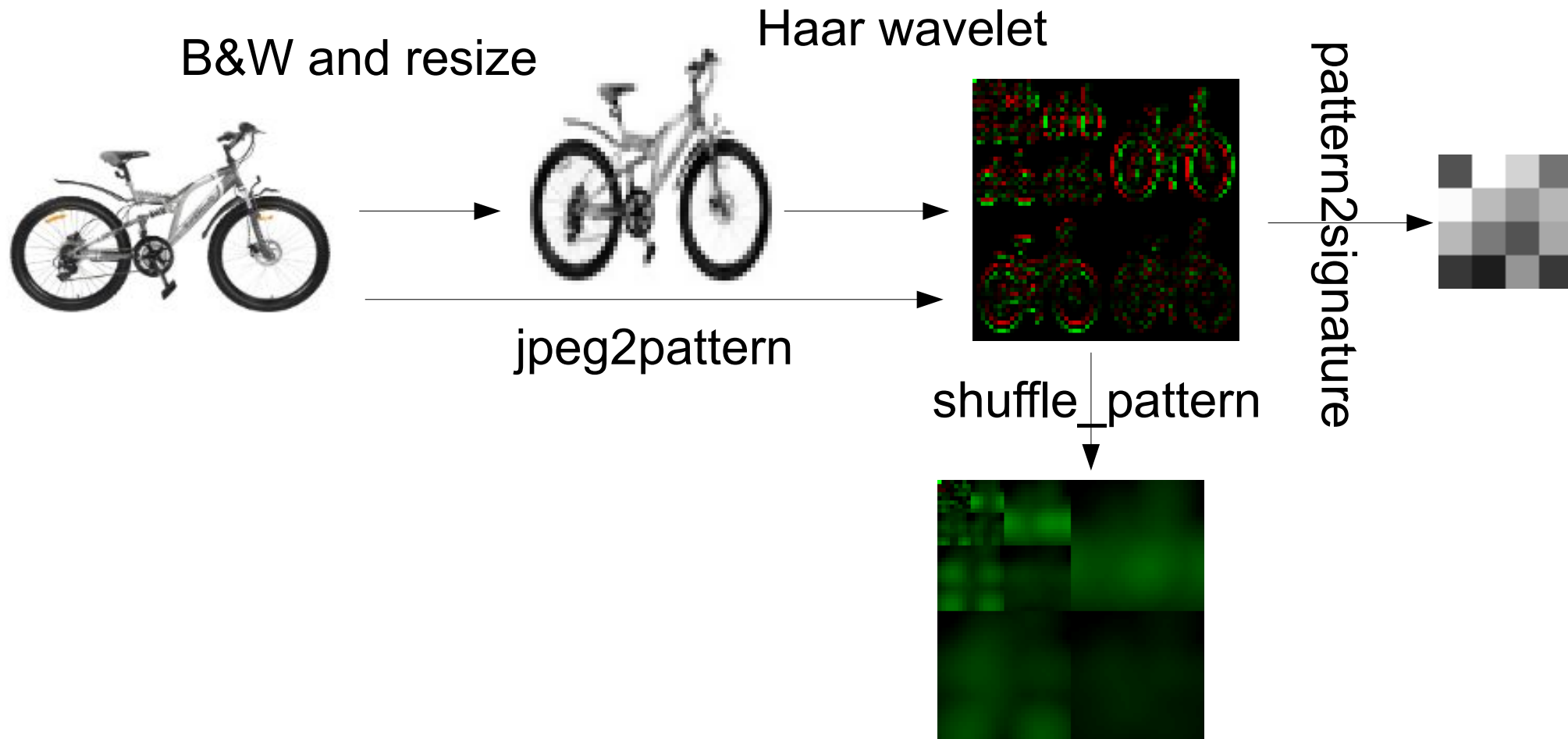
```
SELECT
    id,
    smlr
FROM
(
    SELECT
        id,
        pattern <-> (SELECT pattern FROM pat WHERE id = :id) AS smlr
    FROM pat
    WHERE id <> :id
    ORDER BY
        signature <-> (SELECT signature FROM pat WHERE id = :id)
    LIMIT 100
) x
ORDER BY x.smlr ASC
LIMIT 10
```

Query plan

```
Limit (cost=35.56..35.58 rows=10 width=8) (actual time=23.421..23.423 rows=1)
-> Sort (cost=35.56..35.81 rows=100 width=8) (actual time=23.420..23.422)
    Sort Key: x.smlr
    Sort Method: top-N heapsort  Memory: 25kB
-> Subquery Scan on x (cost=16.54..33.40 rows=100 width=8) (actual time=23.420..23.422)
    -> Limit (cost=16.54..32.40 rows=100 width=100) (actual time=23.420..23.422)
        InitPlan 1 (returns $0)
            -> Index Scan using pat_id_idx on pat (cost=0.00..8.29) (actual time=23.420..23.422)
                Index Cond: (id = 1)
        InitPlan 2 (returns $1)
            -> Index Scan using pat_id_idx on pat (cost=0.00..8.29) (actual time=23.420..23.422)
                Index Cond: (id = 1)
        -> Index Scan using pat_signature_idx on pat (cost=0.00..8.29) (actual time=23.420..23.422)
            Order By: (signature <-> $1)
            Filter: (id <> 1)
            Rows Removed by Filter: 1
Total runtime: 23.465 ms (40082 images)
```

Internals

(bicycle invention)



Thank you for attention!