NON-TEXTUAL DATA EXTRACTION

José Antonio Ruiz Heredia
Joseph Tartivel
Álvaro Honrubia

TABLE OF CONTENTS

01

MOTIVATION

A brief introduction of the project.

02

STATE OF THE ART

Presentation of research paper used for this work.

03

DATASET

Presentation of the dataset used

04

IMPLEMENTATION

Description of the implementation of the project.

05

OTHER EXPERIMENTS

Quick summary of multiple experiments we did during this project.

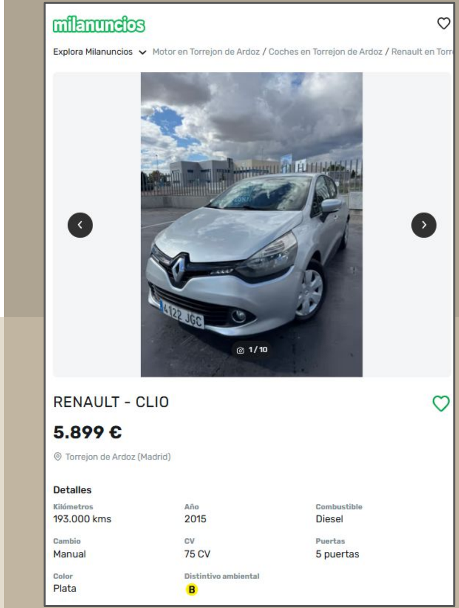
06

IMPROVEMENTS

A view of potentials improvements.

MOTIVATION

- **Problem:** Incomplete and inaccurate car listings on Milanuncios.
- **Solution:** Use CBIR to suggest accurate car details from images.
- **Method:** ORB for shape matching + CCH & SCH for color refinement.
- **Challenges:** Image variations, similar models, and color inconsistencies.
- **Impact:** More accurate listings, better search results, improved user experience.



milanuncios

Explora Milanuncios Motor en Torrejon de Ardoz / Coches en Torrejon de Ardoz / Renault en Torr

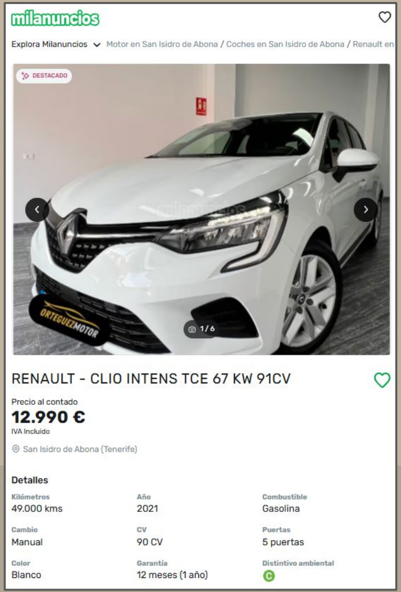
RENAULT - CLIO

5.899 €

Torrejon de Ardoz (Madrid)

Detalles

Kilómetros 193.000 kms	Año 2015	Combustible Diesel
Cambio Manual	CV 75 CV	Puertas 5 puertas
Color Plata	Distintivo ambiental B	



milanuncios

Explora Milanuncios Motor en San Isidro de Abona / Coches en San Isidro de Abona / Renault en

RENAULT - CLIO INTENS TCE 67 KW 91CV

Precio al contado

12.990 €

IVA Incluido

San Isidro de Abona (Tenerife)

Detalles

Kilómetros 49.000 kms	Año 2021	Combustible Gasolina
Cambio Manual	CV 90 CV	Puertas 5 puertas
Color Blanco	Garantía 12 meses (1 año)	Distintivo ambiental C

STATE OF THE ART

ORB: Fast and Robust Feature Detector

Open-source alternative to SIFT & SURF

Key Components:

- **FAST** (with intensity centroid)
- **rBRIEF** (with learning-based feature selection)

Applications:

- **Feature extraction** for tracking & recognition
- Enhanced **object detection**
- Efficient large-scale **image search**
- Hybrid **image retrieval**

COLOR: CCH & SCH

Double Color Histogram for CBIR

- Conventional Colour Histogram (**CCH**) for global color distribution
- Stacked Colour Histogram (**SCH**) to capture texture information

Why It Matters for Car Image Matching

- Robust against *lighting*, *angle*, and *transformation* variations
- Distinguishes cars with similar shapes but different textures/colors
- Improves accuracy in identifying visually similar vehicles

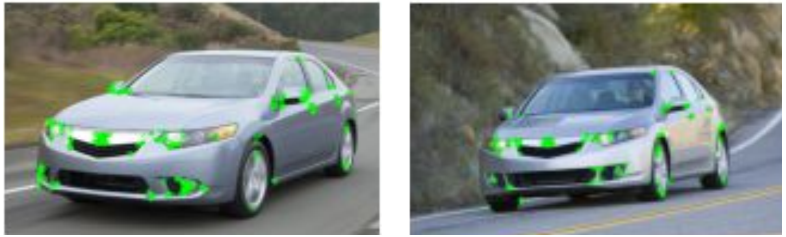
CONSTRUCTING THE TRAINING DATASET

Stanford Cars Dataset with **16,185** images

Diverse range of **vehicle models**

Dataset’s detailed metadata offer additional classification insights for future improvements

IMPLEMENTATION



ORB MATCHER

Uses **ORB (Oriented FAST and Rotated BRIEF)**

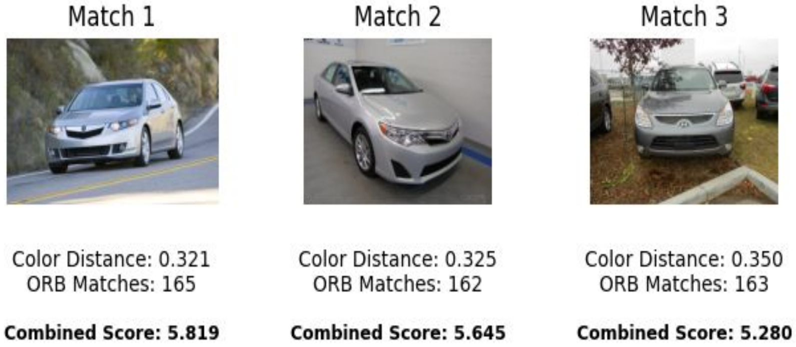
- Compares the **query image's** descriptors against the descriptors of the **dataset**
- Matches based on the number of keypoint using a **brute-force matcher**

COLOR MATCHER

Computes two types of **color-based features**:

- **CCH** based on standard RGB color histograms.
- **SCH** applies blur and averaging over several iterations

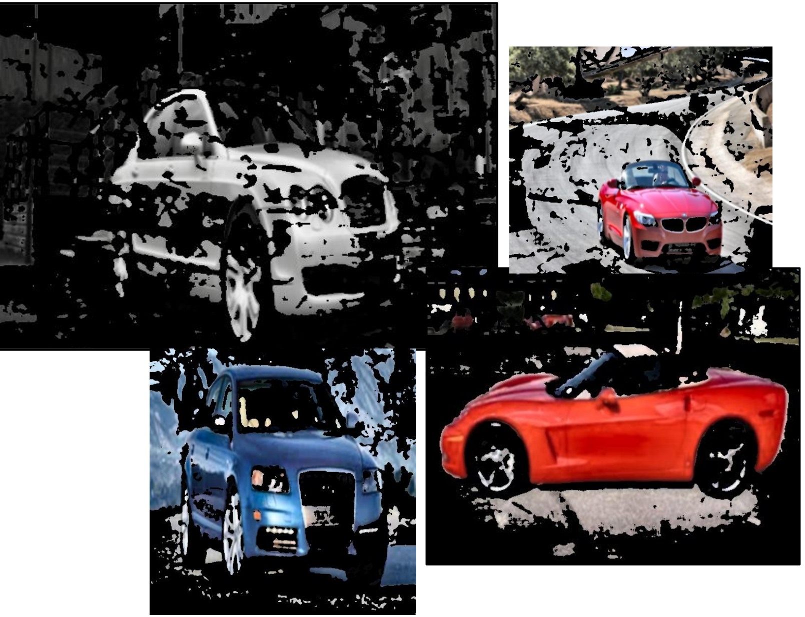
Both combined into a **single vector** to find similar images.



HYBRID MATCHING

integrates the **ORB** and **color-based** results to provide a **combined ranking**

- Filters **top n images** using ORB matcher.
- **Compute** color features of the **filtered list**.
- Calculate a **combined score**
- Display **top results** with metrics and scores.



OTHER EXPERIMENTS

BACKGROUND REMOVAL

Remove the background from images while preserving the foreground

- **Image Standardization**:
- **Background Removal** (BackgroundSubtractorMOG2)
- **Dataset Processing**

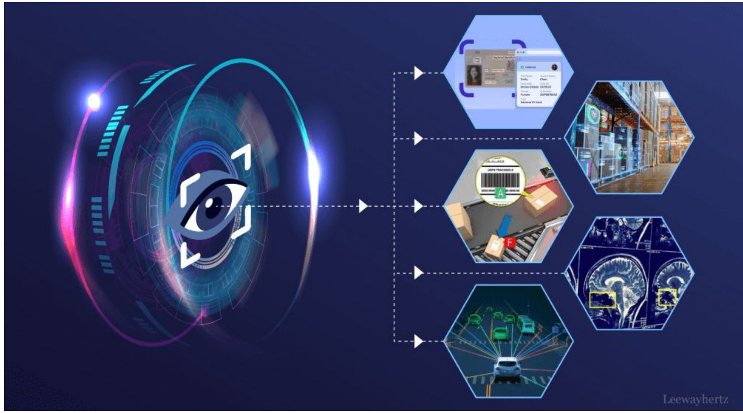
Finally, result didn’t match the required expectations

ORB MATCHER ALTERNATIVES

Experimented with other two feature matching techniques:

- **FLANN** uses Locality Sensitive Hashing (**LSH**) for fast nearest neighbor searches.
- **HNSW** is a graph-based algorithm from **Faiss** that provides efficient similarity search for high-dimensional data.

Both methods didn’t enhance the outputs by **BF Matcher**



Leewayhertz

IMPROVEMENTS

Background removal

Implement **robust segmentation techniques** to isolate subject from surroundings
Evaluate performance of **different algorithms** across varied conditions

Standardize images

Develop **normalization** procedures for lighting variations
Create **position calibration** for consistent vehicle orientation
Implement **bias correction** methods for environmental factors

Experiment ORB matcher alternatives

Evaluate ***SIFT***, ***SURF***, ***AKAZE*** and other feature detection algorithms
Compare **performance metrics**: accuracy, processing time, robustness

Methods beyond color matcher

Investigate **texture-based** recognition approaches
Explore **deep learning-based** feature extraction
Research **hybrid methods** combining multiple recognition strategies