

Dieter Janzen & Bernd Ackermann

TWO
EYES
SEE

MORE

Using machine learning in the first steps
of generating 3D models from 2D images

Source: Photography by Joseph L. Bates (~1870)



TABLE OF CONTENTS



01 THE CHALLENGE

How to infer 3D models from photos

03 THE MODEL

What is a LoFTR and can you eat it?



02 THE DATA

EDA and obstacles of the dataset



04 THE FINAL PRODUCT

An interactive LoFTR dashboard





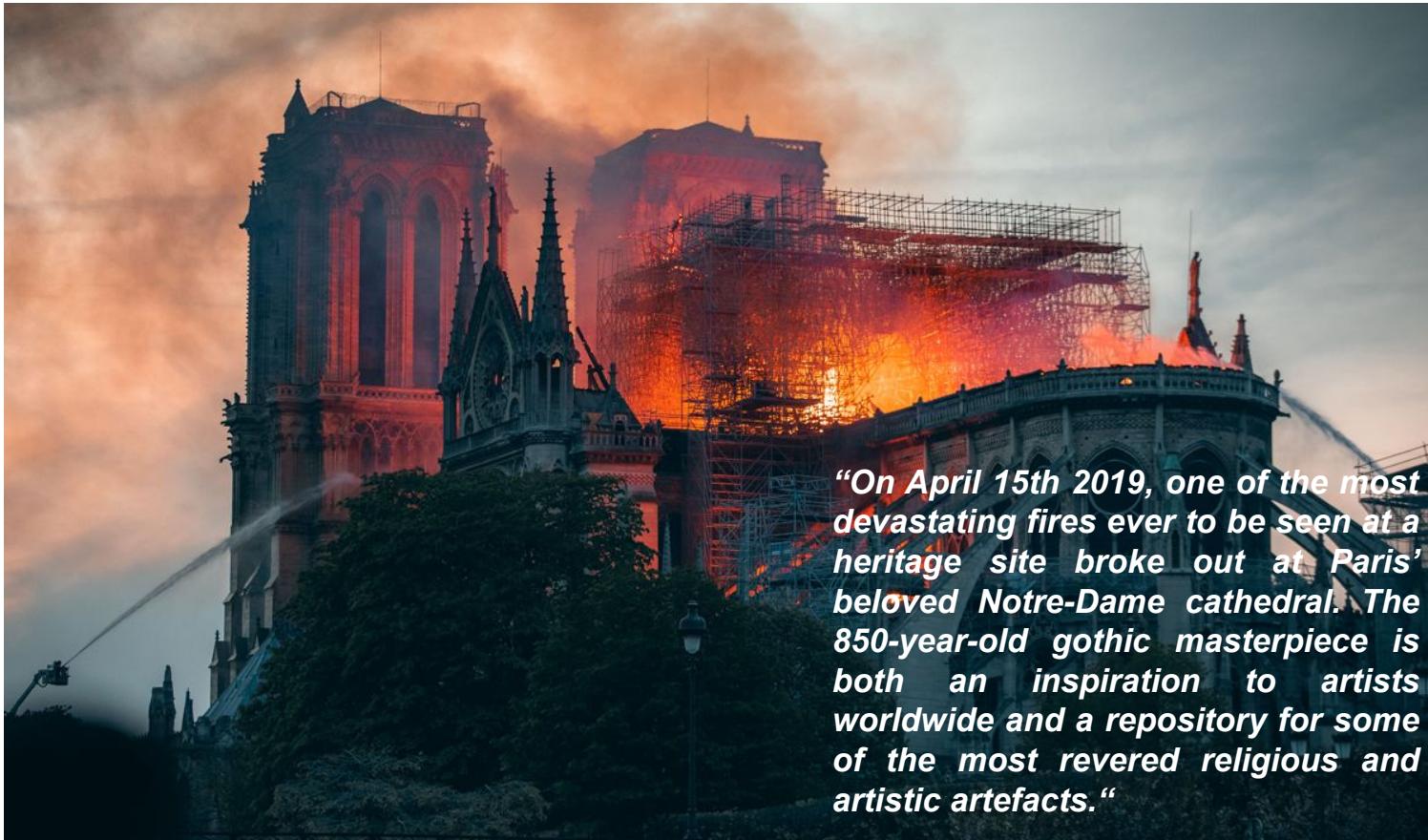
01

THE CHALLENGE



How to infer 3D models from photos





“On April 15th 2019, one of the most devastating fires ever to be seen at a heritage site broke out at Paris’ beloved Notre-Dame cathedral. The 850-year-old gothic masterpiece is both an inspiration to artists worldwide and a repository for some of the most revered religious and artistic artefacts.”



Andrew J. Tallon

* 12 March 1969

† 16 November 2018

Source: Cristina Abellán Matamoros. Unique 3D model of Notre Dame cathedral could help reconstruction efforts. *euronews*



HOW TO MAKE 3D RECONSTRUCTIONS



STRUCTURE-FROM-MOTION (SfM)

- Reconstruction of 3D models from images
- Usually **requires very controlled conditions** (i.e. a “calibrated” camera)
- Images of a sequence are compared, respective to their **relative positions and rotations**



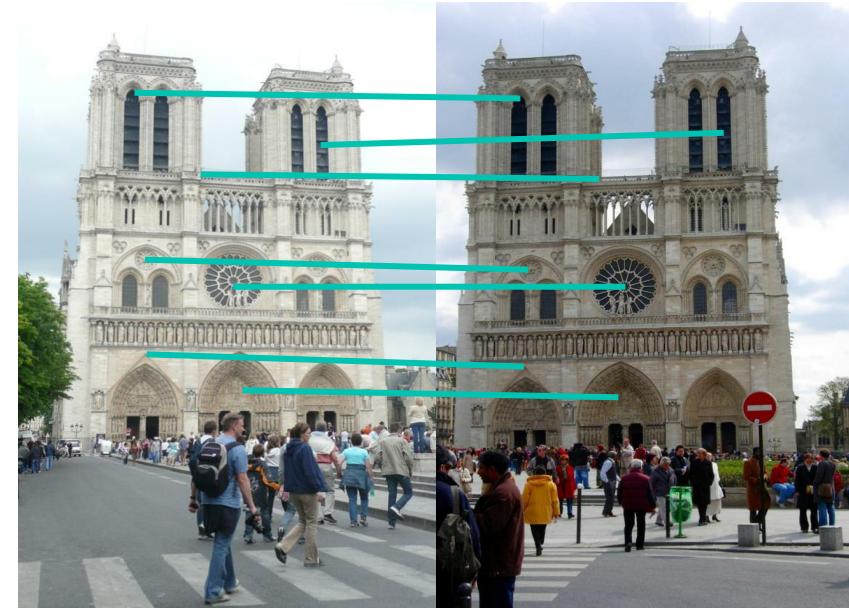
SfM from uncalibrated images would help monument conservationists immensely





IMAGE MATCHING

Finding **common features** among two images allows us to calculate a **Fundamental Matrix (FM)**, a metric used to estimate the relative position/rotation of two cameras.



FUNDAMENTAL MATRIX (FM)

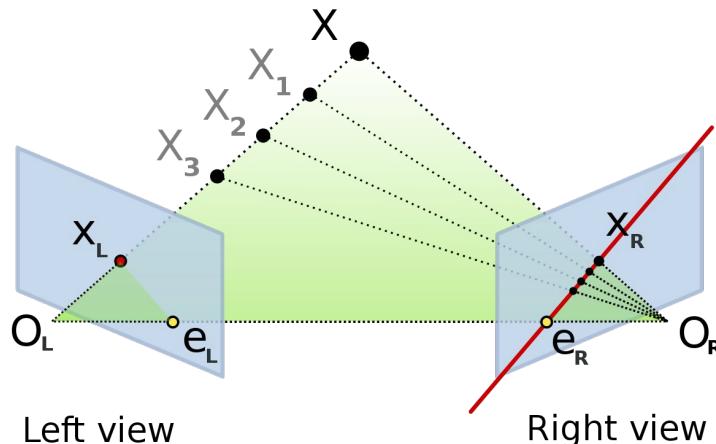


TARGET: FINDING THE FM

By identifying (at least) 8 pairs of points $X_L|X_R$, found in both images, the **FM** can be calculated

RELATIVE IMAGE POSITIONS

Using the **FM**, both camera's *focal lengths*, and *principal point offsets* allows us to estimate the relative camera positions and rotations.





02

THE DATA

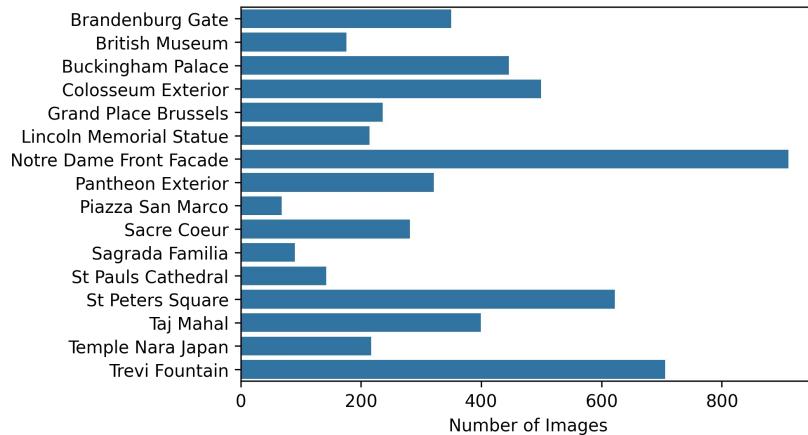


EDA and obstacles of the dataset

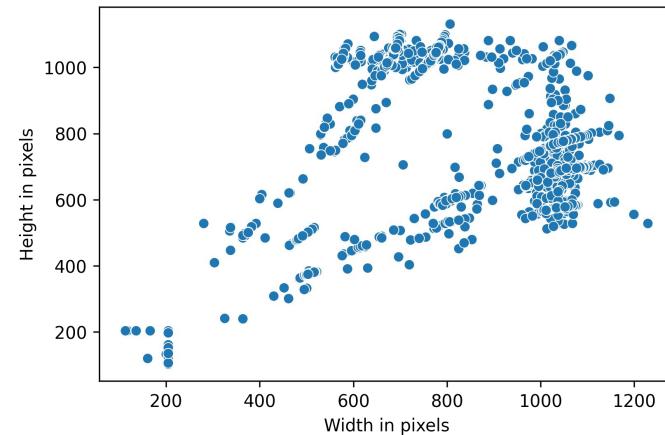


DATA OVERVIEW

5678 images, distributed unevenly over 16 scenes



Images come in various pixel sizes and aspect ratios



DATA OVERVIEW

Brandenburg Gate
1075 x 787 px



British Museum
1041 x 686 px



Buckingham Palace
1077 x 583 px



Colosseum Exterior
1046 x 776 px



Grand Place Brussels
1072 x 786 px



Lincoln Memorial Statue
789 x 1080 px



Notre Dame Front Facade
1038 x 723 px



Pantheon Exterior
764 x 1025 px



Piazza San Marco
1045 x 776 px



Sacre Coeur
768 x 1025 px



Sagrada Familia
688 x 1053 px



St Pauls Cathedral
1077 x 696 px



St Peters Square
1039 x 688 px



Taj Mahal
1001 x 750 px



Temple Nara Japan
1044 x 686 px



Trevi Fountain
1022 x 684 px



Images not only show the target scene, but also people, obstructed views, custom filters, etc.

IMAGE PAIRS



- In the provided training data, every image (5678 total) was matched with all other images of the same scene
 - This results in **1,416,814** image pairs
- Every pair has a covisibility score between 0 and 1, where 1 means it's a perfect match

Good covisibility (0.713)



Bad covisibility (0.007)



CAMERA CALIBRATION

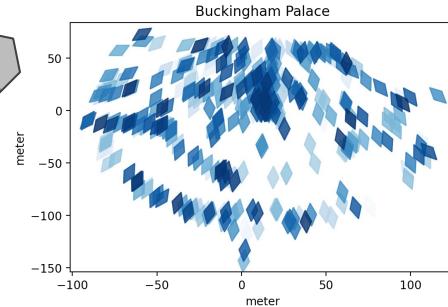
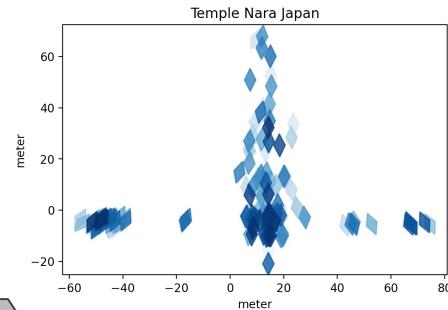
FUNDAMENTAL MATRIX
+
CAMERA INTRINSICS



CAMERA EXTRINSICS



SCALING FACTORS





03

THE MODEL



LoFTR: The Fellowship of the
Fundamental Matrix





LoFTR



Simpler ML models (e.g. SVR) failed to predict the fundamental matrix

LoFTR: Local Feature Matching with Transformers (Sun et al., CVPR 2021)

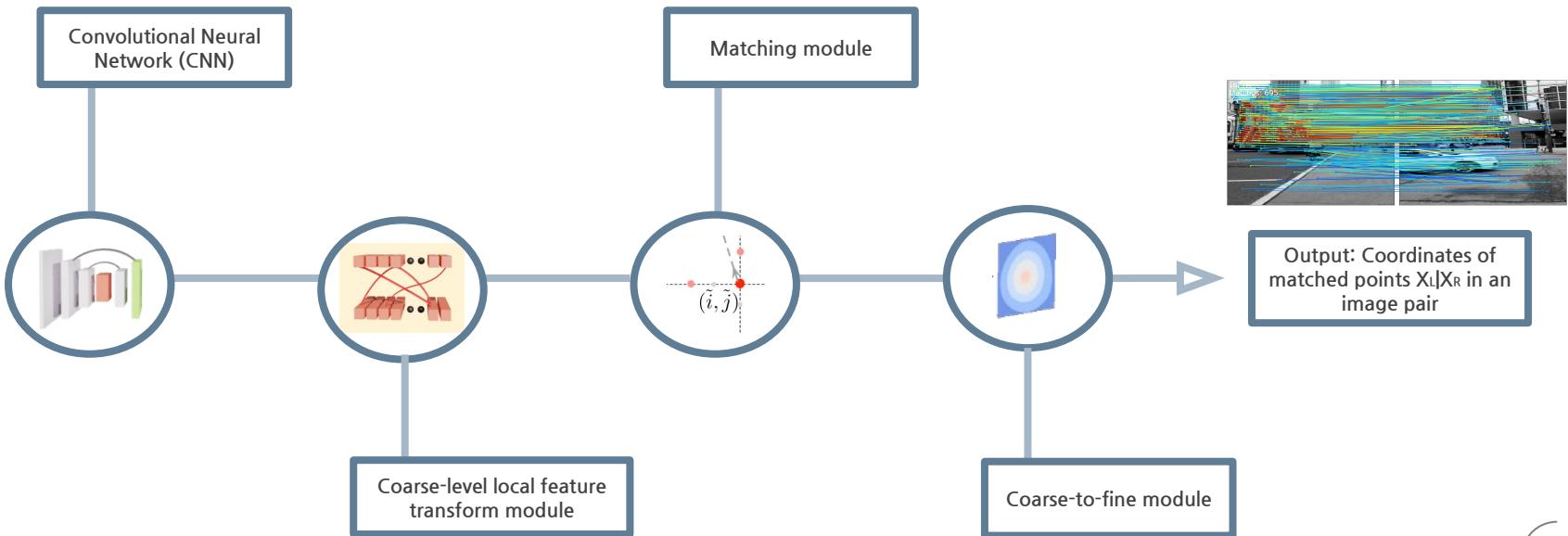




Image Matching - By Hand

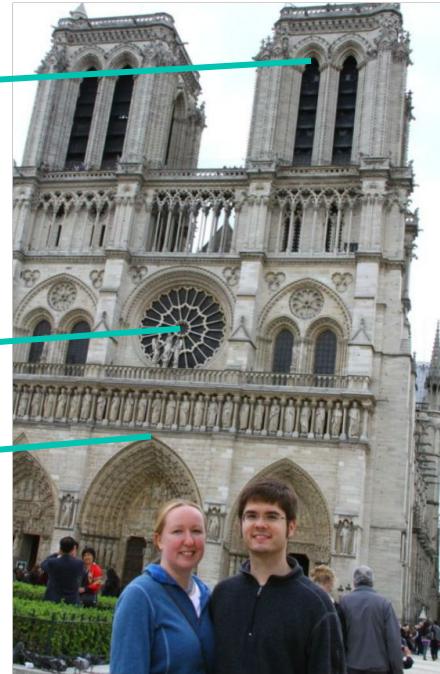
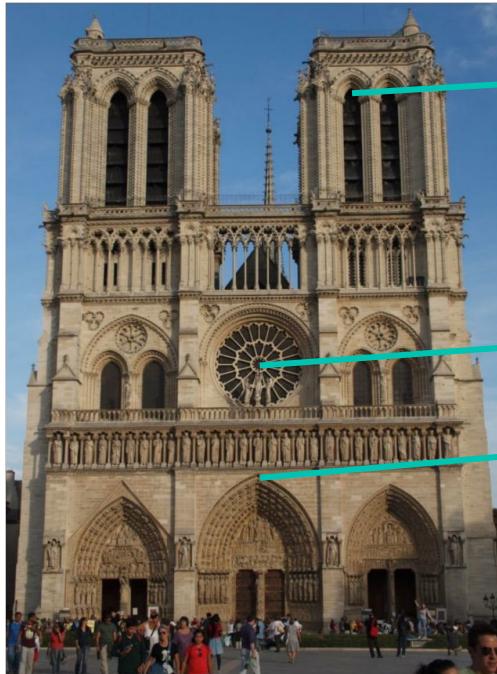
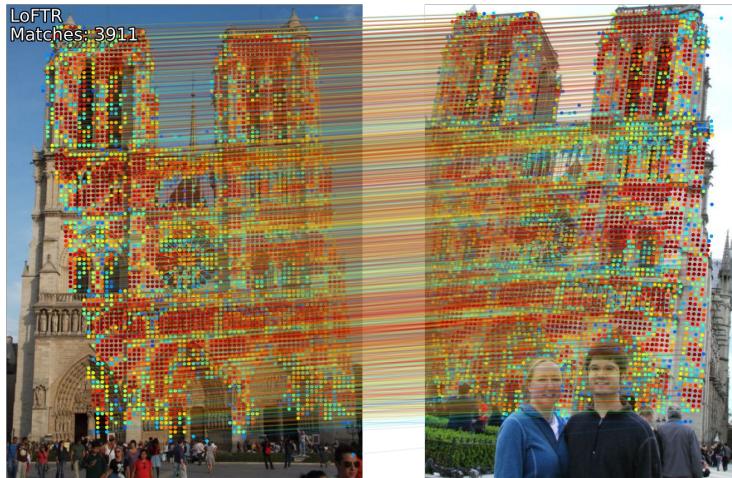


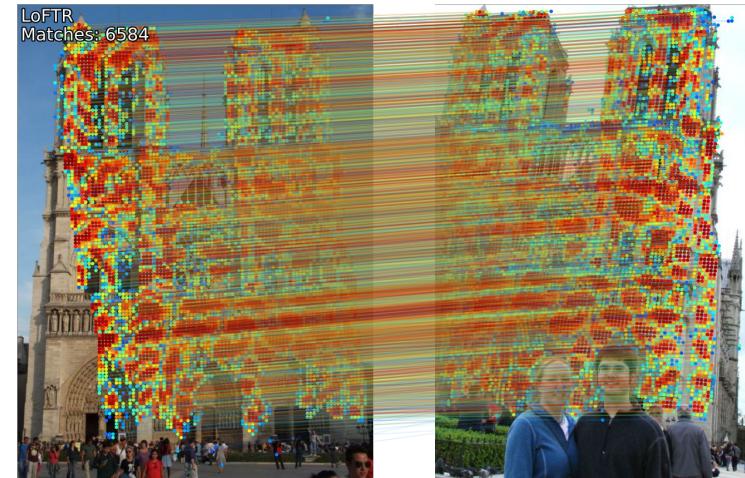
Image Matching - LoFTR

Baseline



Accuracy of calculated
camera position and
rotation: 60%

Optimized



Accuracy: 90%
(positional error: < 20 cm,
rotational error: < 2°)





04

THE FINAL PRODUCT



An interactive LoFTR dashboard



DASHBOARD FEATURES



INTERACTIVE DATA VIZ

Images can be selected based on position, previewed and paired



LoFTR PREDICTION

Show image point pairings as basis for the FM calculation and predicted camera positions

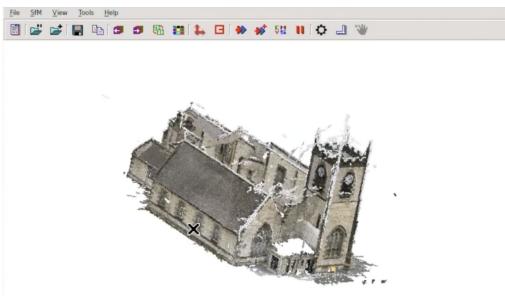
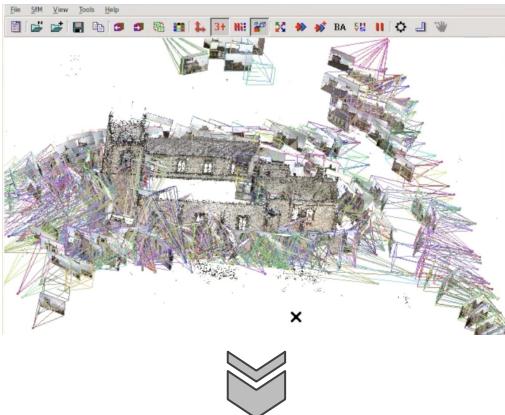


UPLOAD AND MATCH OWN IMAGES

Upload your own images and match them using LoFTR



FUTURE WORK



VisualSfM

FURTHER OPTIMIZATION

The accuracy of our LoFTR model could be further increased by using it in an ensemble with other image matchers

3D RECONSTRUCTION

Existing 3D reconstruction tools, like VisualSfM, could be improved by implementing LoFTR



Improving monument conservation efforts



THANKS!

Do you have any questions?



- • • CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#)