Matchmove

Multimedia Techniques & Applications

Yu-Ting Wu

(with slides borrowed from Prof. Yung-Yu Chuang)

1

Jurassic Park (1993)

2

How to Composite Virtual and Real?

- In the real world, we use a **camera** to record the information of the real scene
- In a virtual world, we use a **virtual camera** to record the information of the virtual scene
- **Idea:** make the virtual camera **sync** with the real-world camera and **put the virtual objects in the right places**

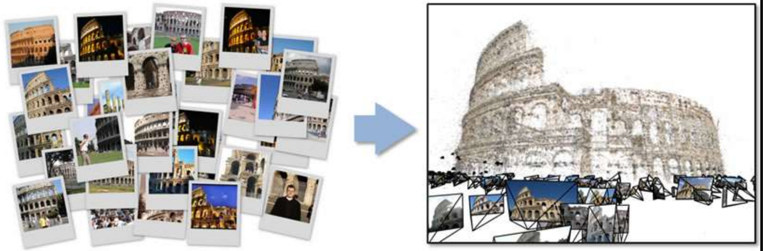
3

How to Composite Virtual and Real? (cont.)

4

Matchmove (Structure from Motion)

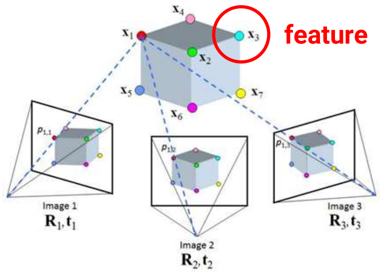
- **Structure from Motion:** automatic recovery of camera motion and scene structure from two or more images
- Also called **matchmove** in film production



Multimedia Techniques and Applications 2022

Matchmove (Structure from Motion)

- **Structure from Motion:** automatic recovery of camera motion and scene structure from two or more images
- Also called **matchmove** in film production



Multimedia Techniques and Applications 2022

Features

- Also known as **interesting points**, **salient points**, or **keypoints**
- Points that you can easily point out their **correspondences** in **multiple images** using only **local information**

Multimedia Techniques and Applications 2022

Desired Properties for Features

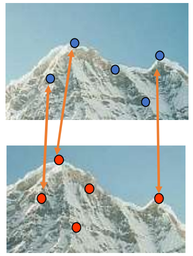
- **Distinctive**
  - A single feature can be correctly matched with high probability
- **Invariant**
  - Invariant to scale, rotation, illumination and noise for robust matching across a substantial range of distortion, viewpoint change and so on

Applications

- **Object or scene recognition**
- **Matchmove (structure from motion)**
- **Stereo**
- **Motion tracking**
- **…**

Components

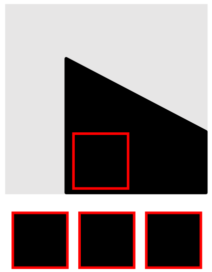
- **Feature detection** locates where they are
- **Feature description** describes what they are
- **Feature matching** decides whether two are the same one



Moravec Corner Detector

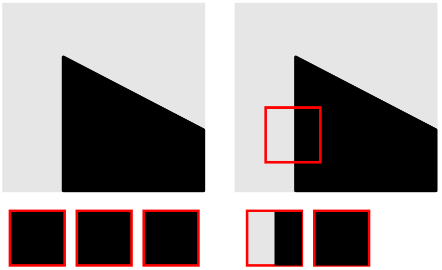
- We should easily recognize the point by looking through a **small window**
- Shifting a window in any direction should give a large change in **intensity**

Moravec Corner Detector (cont.)



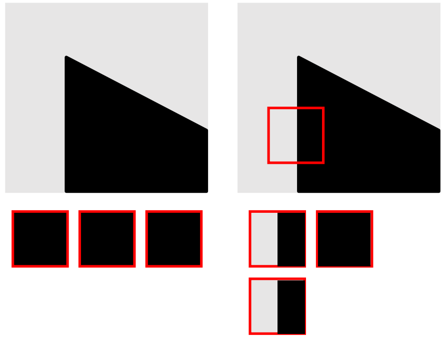
flat

Moravec Corner Detector (cont.)



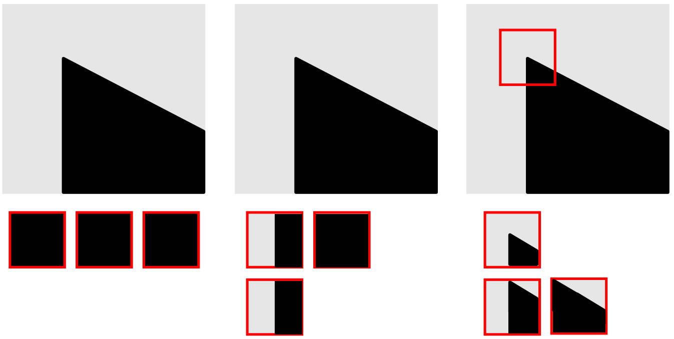
flat

Moravec Corner Detector (cont.)



flat edge

Moravec Corner Detector (cont.)



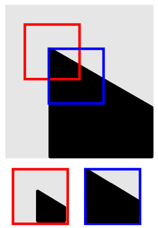
flat edge corner isolated point

Moravec Corner Detector (cont.)

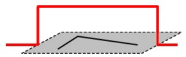
- Change of intensity for the shift [u, v]

$$E(u,v) = \sum_{x,y} w(x,y)\left[I(x+u, y+v) - I(x,y)\right]^2$$

window function — shifted intensity — intensity



Window function $w(x,y)$ =



1 in window, 0 outside

Four shifts: (u,v) = (1,0), (1,1), (0,1), (-1, 1)
Look for local maxima in *min{E}*

Problems of Moravec Detector

- Noisy response due to a binary window function
- Only a set of shifts at every 45 degree is considered
- Only minimum of E is taken into account

➔ **Harris corner detector solves these problems**

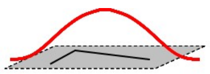
Harris Corner Detector

- Noisy response due to a binary window function

➔ **Use a Gaussian function**

$$w(x,y) = \exp\left(-\frac{(x^2+y^2)}{2\sigma^2}\right)$$

Window function $w(x,y)$ =



Gaussian

Harris Corner Detector (cont.)

- Only a set of shifts at every 45 degree is considered

➔ **Consider all small shifts by Taylor's expansion**

$$E(u,v) = \sum_{x,y} w(x,y)\left[I(x+u, y+v) - I(x,y)\right]^2$$
$$= \sum_{x,y} w(x,y)\left[I_x u + I_y v + O(u^2, v^2)\right]^2$$

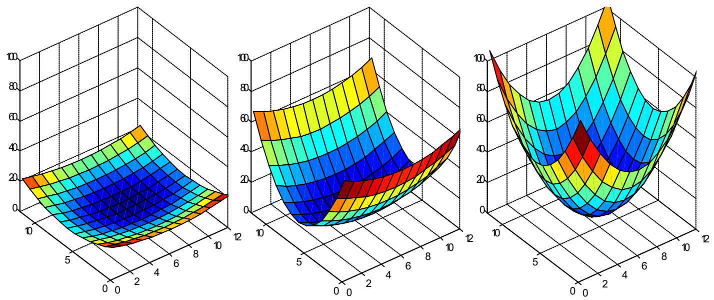
$$E(u,v) = Au^2 + 2Cuv + Bv^2$$

We can obtain a new measurement by investigating the shape of the error function

$$A = \sum_{x,y} w(x,y) I_x^2(x,y)$$
$$B = \sum_{x,y} w(x,y) I_y^2(x,y)$$
$$C = \sum_{x,y} w(x,y) I_x(x,y) I_y(x,y)$$

Harris Corner Detector (cont.)

- High-level idea: what shape of the error function will we prefer for features?

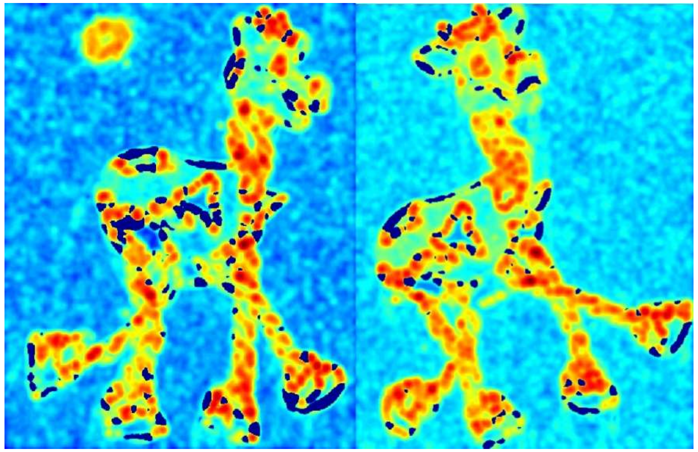


flat edge corner

Harris Corner Detector (Input)

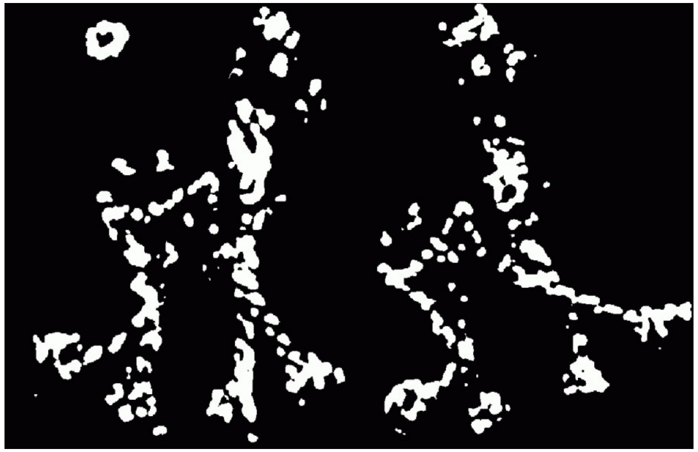
21

Harris Corner Detector (Response)



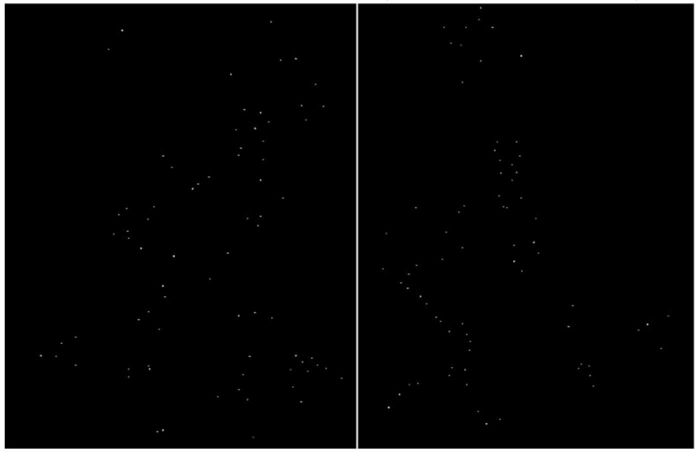
22

Harris Corner Detector (Theshold)



23

Harris Corner Detector (Local Maximum)

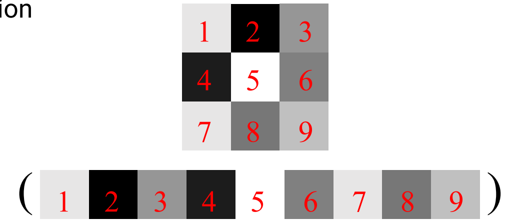


24

Harris Corner Detector (Output)

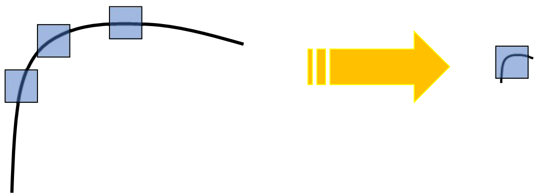
Feature Description

- Now we know where the features are
- But how to match them?
- What is the descriptor for a feature? The simplest solution is the intensities of its spatial neighbors
- This might not be robust to brightness change or small shift/rotation



Problems of Harris Detector

- Not invariant to image scale

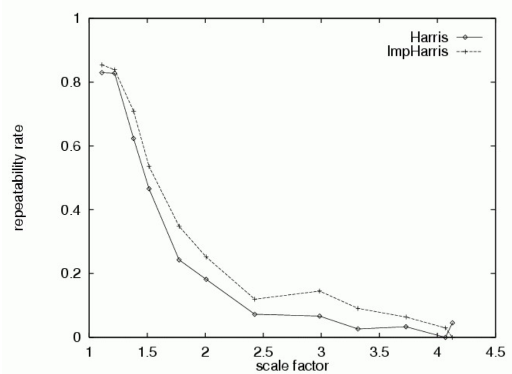


All points will be classified as edges

Corner !

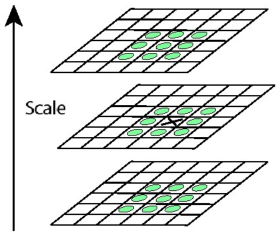
Problems of Harris Detector (cont.)

- Not invariant to image scale



SIFT

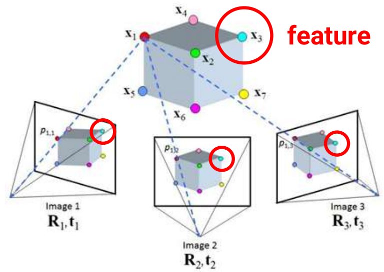
- Stands for **Scaled Invariant Feature Transform**
- For **scale invariance**, search for stable features **across all possible scales** using a continuous function of scale, scale space.



29

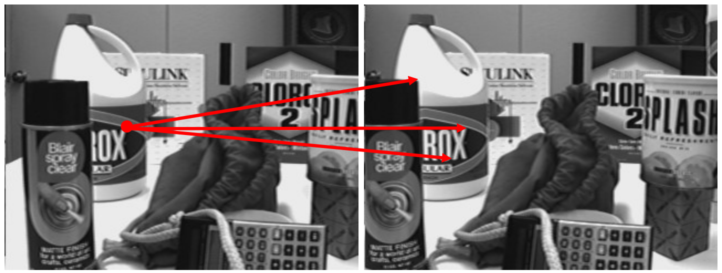
Tracking

- If we detect a feature point in one frame, how do we keep tracks of it in other frames?



30

Tracking (cont.)



31

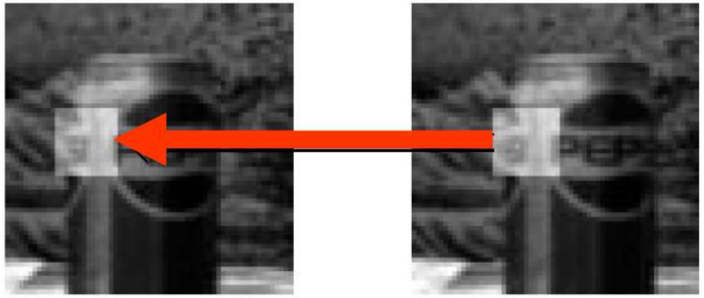
Three Assumptions of Tracking

- **Brightness consistency**
- **Spatial coherence**
- **Temporal persistence**

32

Brightness Consistency

- Image measurement (e.g. brightness) in a small region remain the same although their location may change

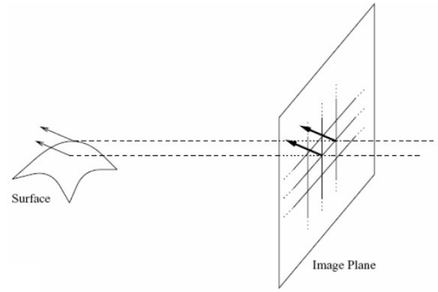


33

Multimedia Techniques and Applications 2022

Spatial Coherence

- Neighboring points in the scene typically belong to the same surface and hence typically have similar motions.
- Since they also project to nearby pixels in the image, we expect spatial coherence in image flow.

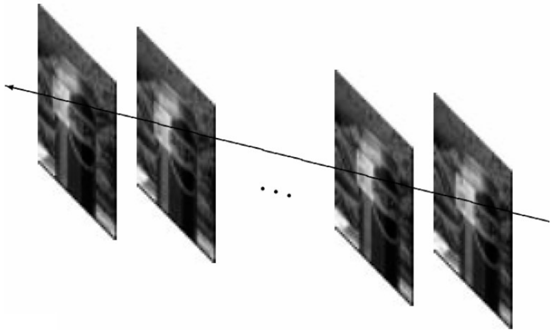


34

Multimedia Techniques and Applications 2022

Temporal Persistence

- The image motion of a surface patch changes gradually over time



35

Multimedia Techniques and Applications 2022

Simple Tracking Approach

- Minimize brightness difference

$$E(u,v) = \sum_{x,y} \left(I(x+u, y+v) - T(x,y) \right)^2$$

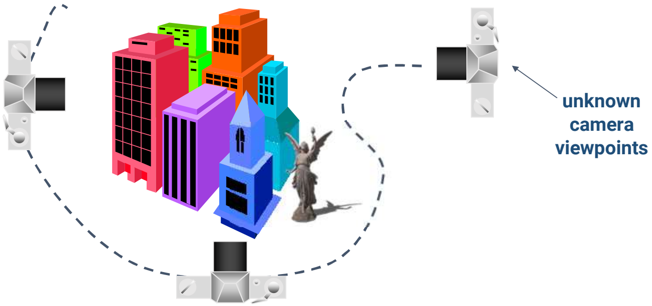
- For each offset (u, v) compute E(u, v)
- Choose (u, v) which minimizes E(u, v)
- Problems:
  - Not efficient
  - Only sub-pixel accuracy

There are more efficient algorithms (e.g. Lucas-Kanade) for tracking

36

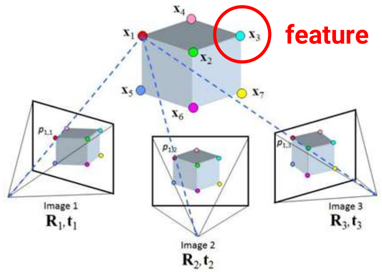
Back to the Matchmove Problem

We need to reconstruct the **camera path**
We also need to reconstruct the **(partial) scene geometry**



Back to the Matchmove Problem (cont.)

- For the scene geometry, we only recover the 3D position of **feature points**

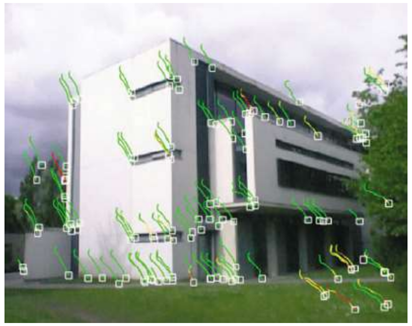


Matchmove Pipeline

2D feature tracking → **3D estimation** → **optimization (bundle adjust)** → **geometry fitting**

2D Feature Tracking

- Detect good features (e.g. by SIFT)
- Find correspondences between frames



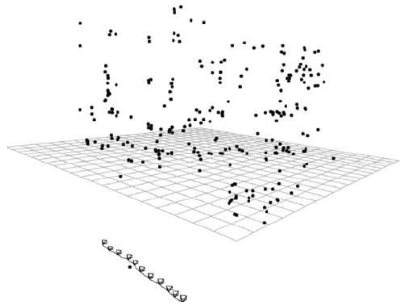
3D Estimation

- Use 2 or 3 views at a time
- Solve an optimization problem

41

Optimization

- Iterative refine estimates



42

Geometry Fitting

- Recover surface by image-based triangulation, silhouettes, or stereo

43

Matchmove in Blender

44

Steps

- **Prepare a video**
- **Extract image sequence (optional)**
- **3D estimation: solve and optimize camera motion and scene geometry**
- **Import 3D models and edit their animations**
- **Output video**

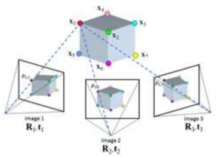
45

Prepare a Video

46

Prepare a Video

- You can either capture your video (suggested) or download ones from the internet
- **Some useful tips**
  - It is better to have **many features** in your video
    - And the features should exist in the entire video (**especially for the ground**)
  - Not too long (if it is, subdivide it and edit each part separately)
  - Your camera should have both **translation** and **rotation**
  - Your video should have **large parallax**



47

Bad example

48

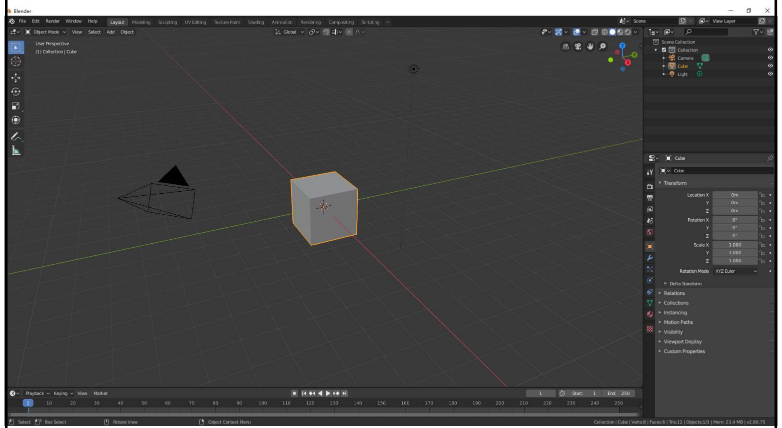
Good example

Extract Image Sequence (Optional)

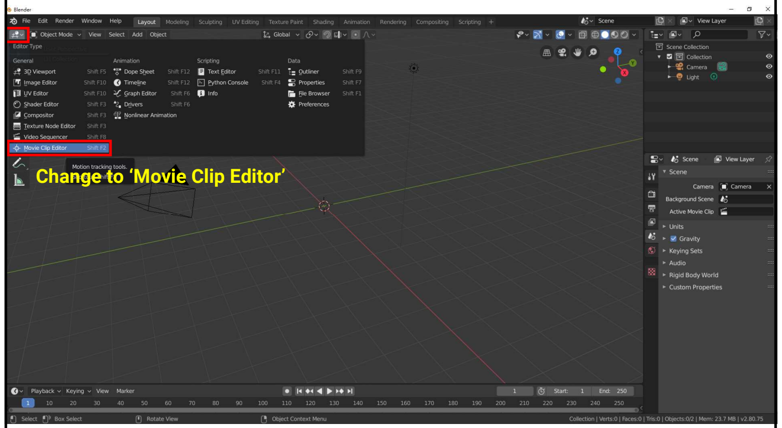
Convert a Video into Image Sequence

- **Why?**
  - Video files have compression built-in
  - Using image sequence leads to better run-time performance

Delete unused objects

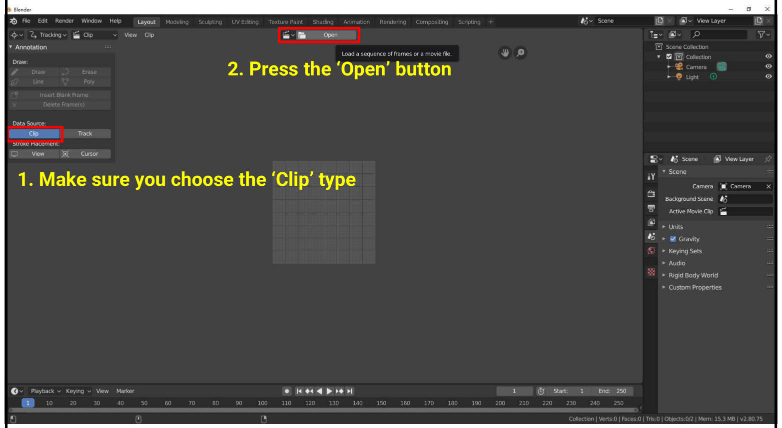


Load Your Video (1)



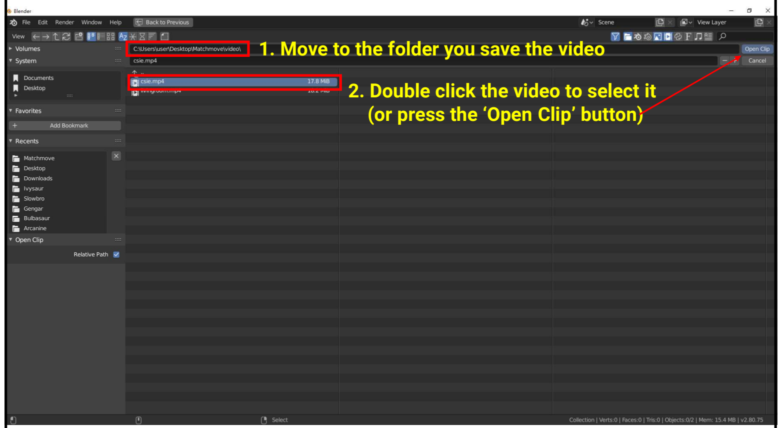
53

Load Your Video (2)



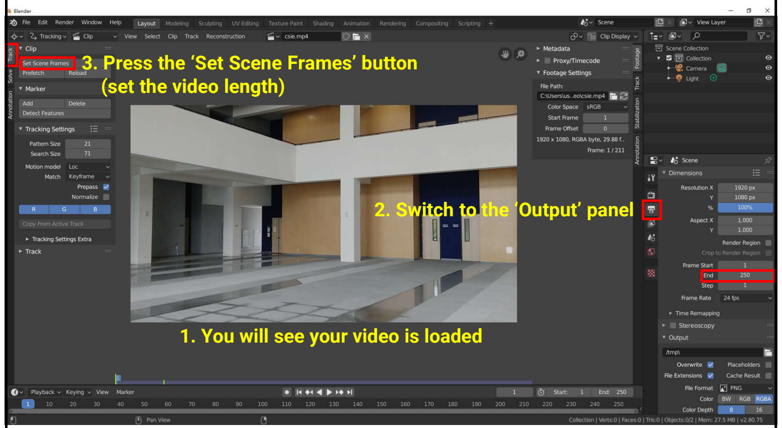
54

Load Your Video (3)



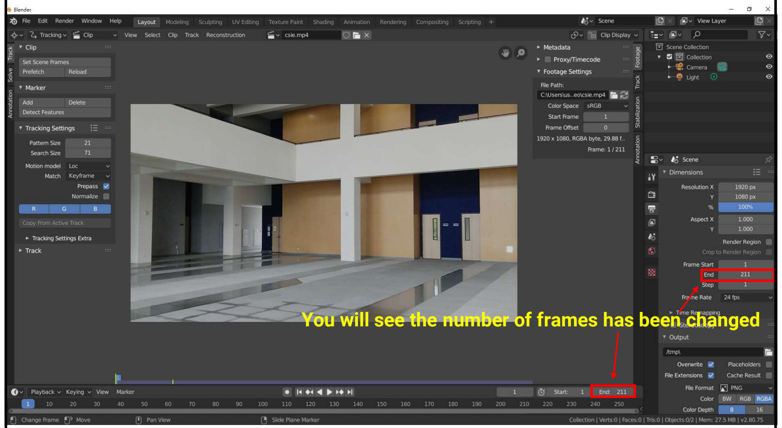
55

Load Your Video (4)



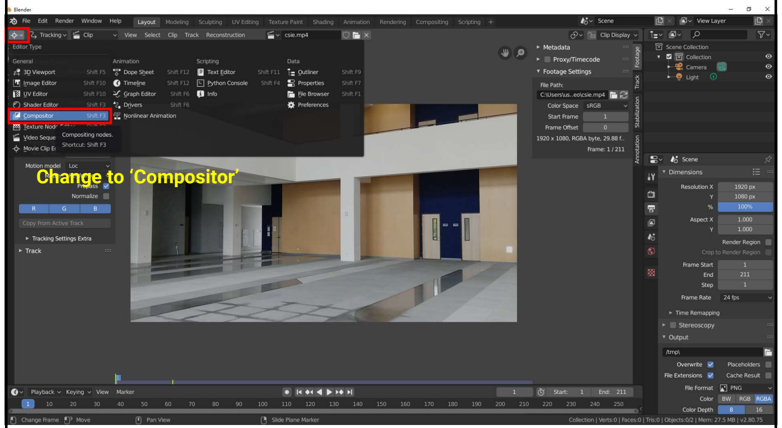
56

Load Your Video (5)



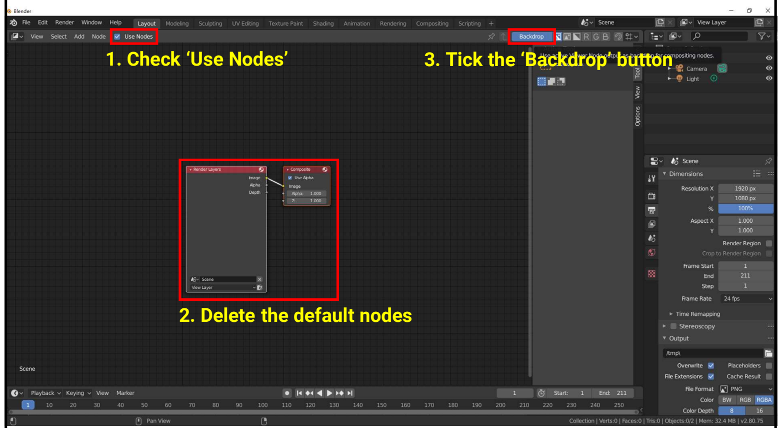
57

Generate Image Sequence (1)



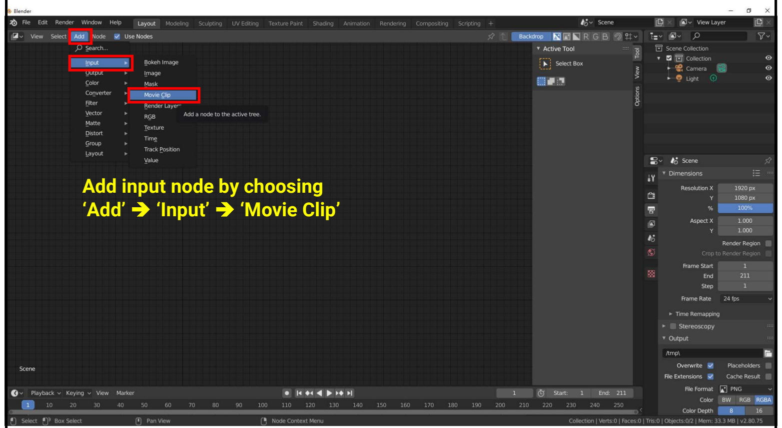
58

Generate Image Sequence (2)



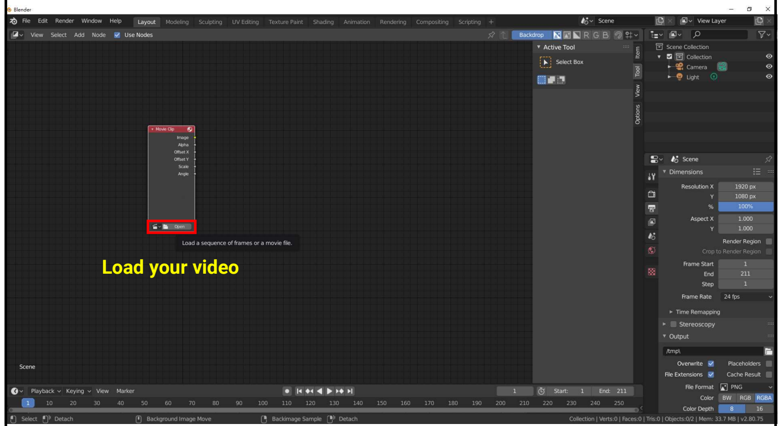
59

Generate Image Sequence (3)



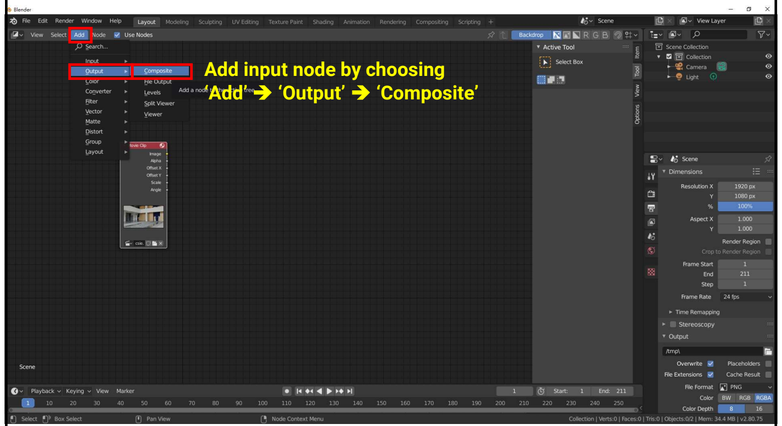
60

Generate Image Sequence (4)



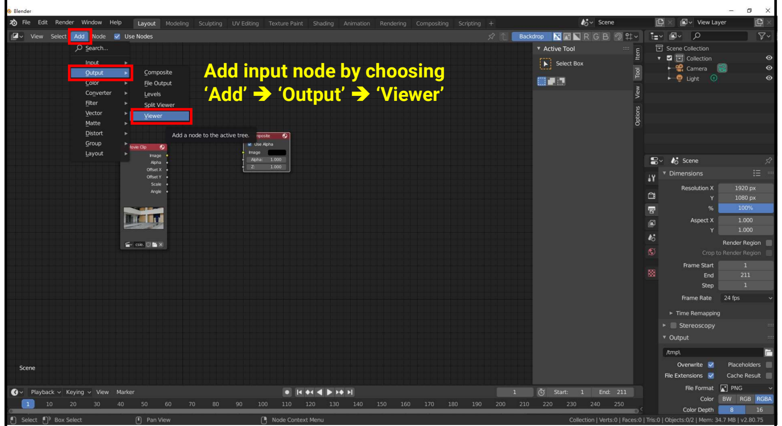
61

Generate Image Sequence (5)



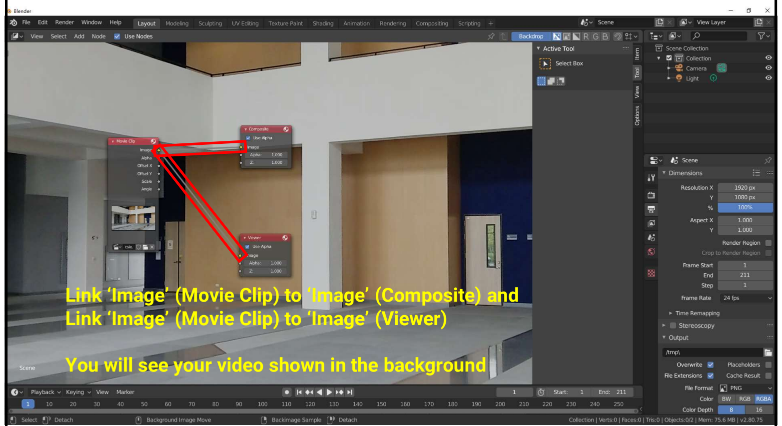
62

Generate Image Sequence (6)



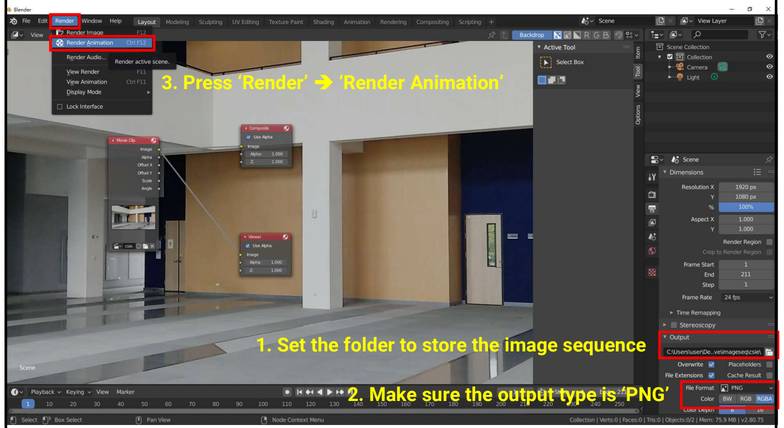
63

Generate Image Sequence (7)



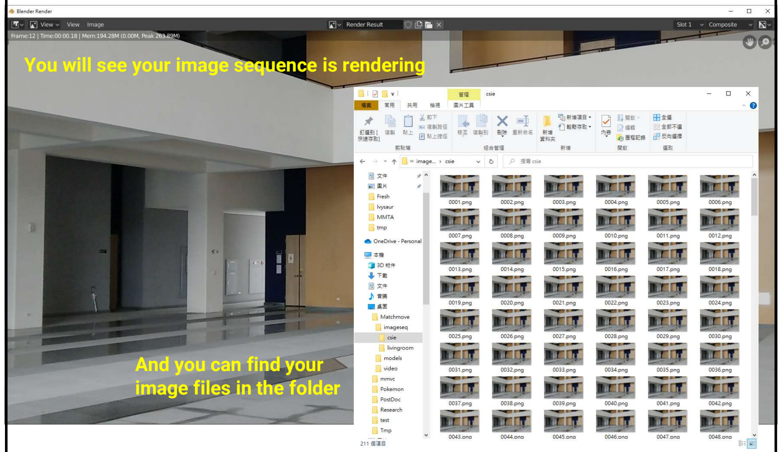
64

Generate Image Sequence (8)



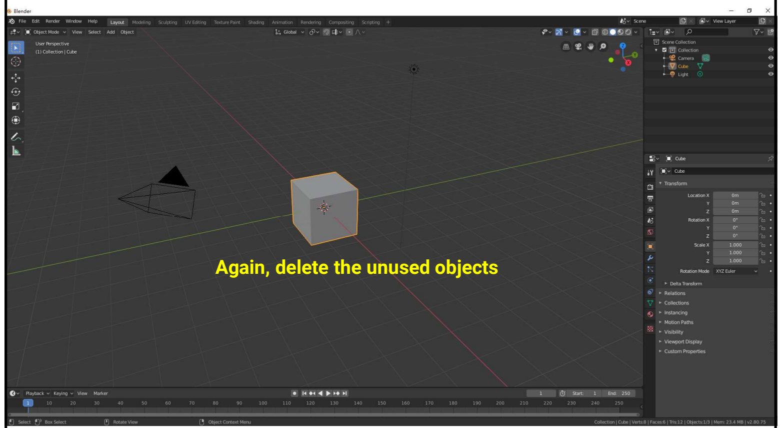
65

Generate Image Sequence (9)



66

Close your Blender and Reopen It

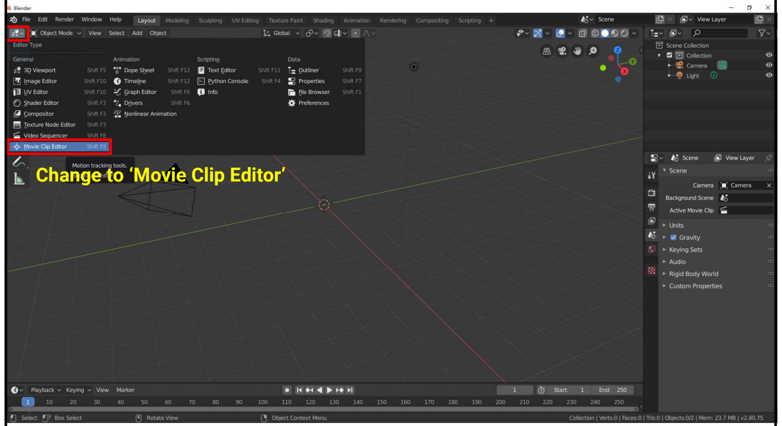


67

3D Estimation

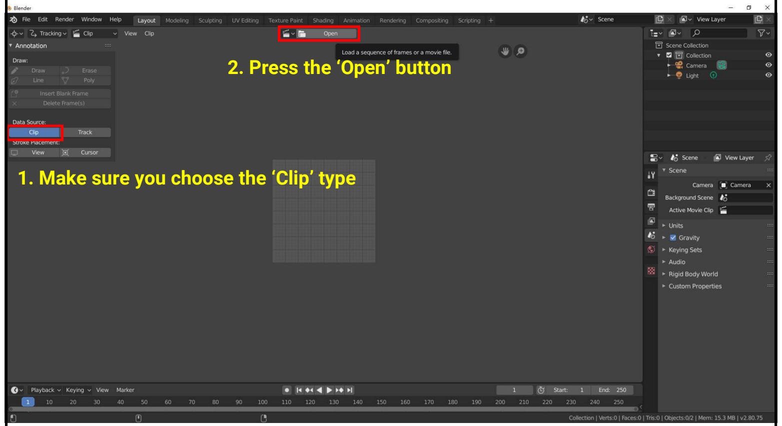
68

Load the Image Sequence (1)



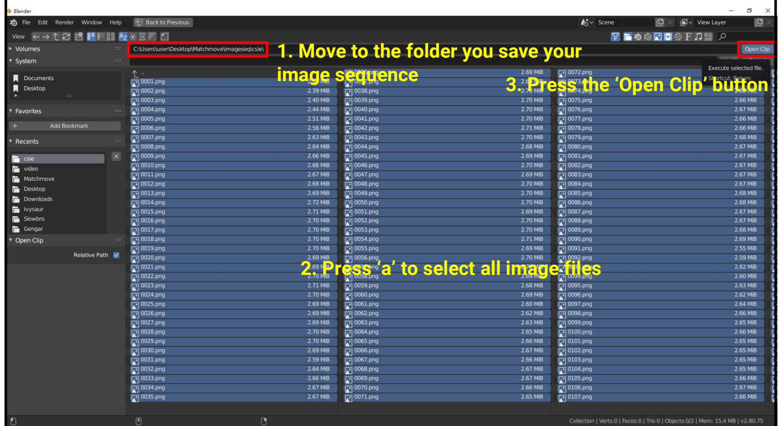
69

Load the Image Sequence (2)



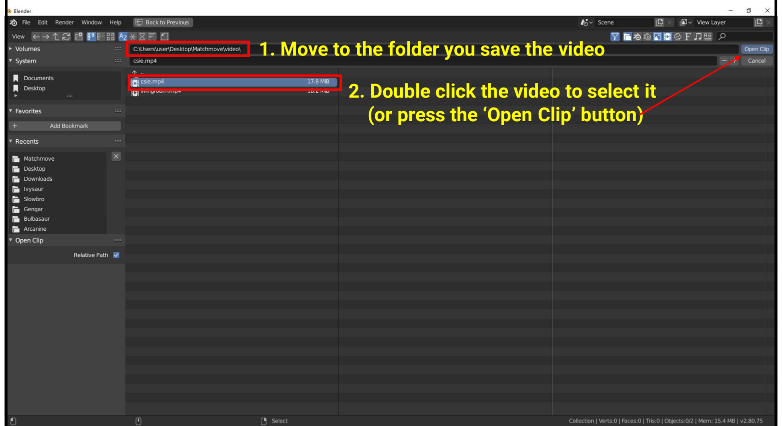
70

Load the Image Sequence (3)



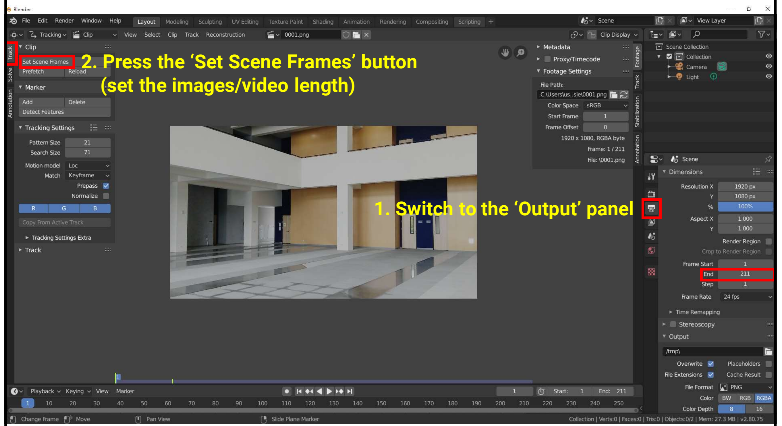
71

Or Load Your Video (If you skip step 1)

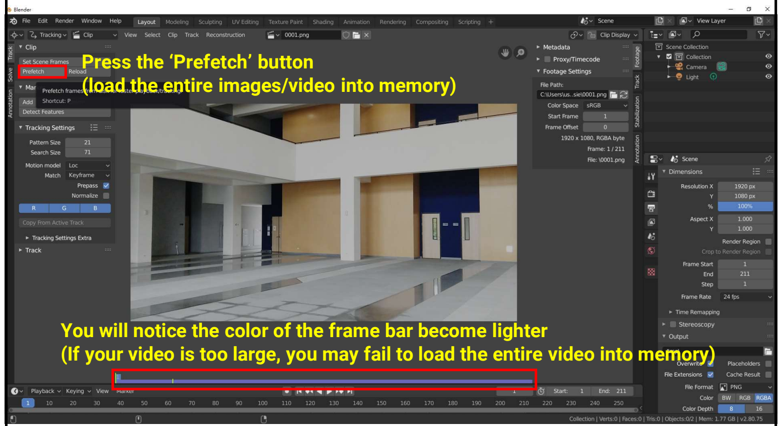


72

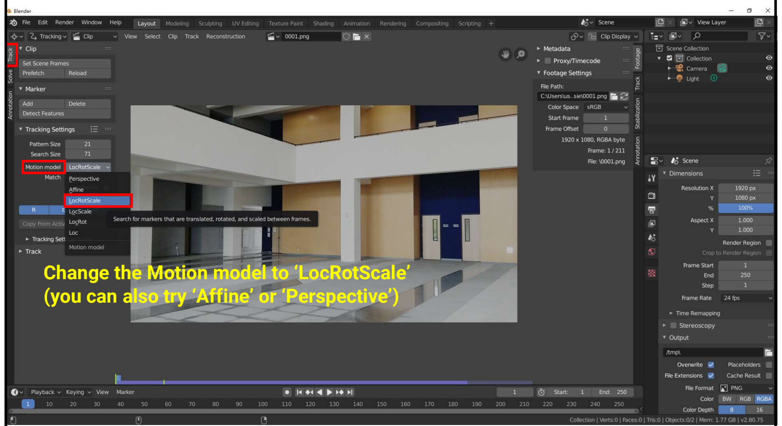
Set Input Data (1)



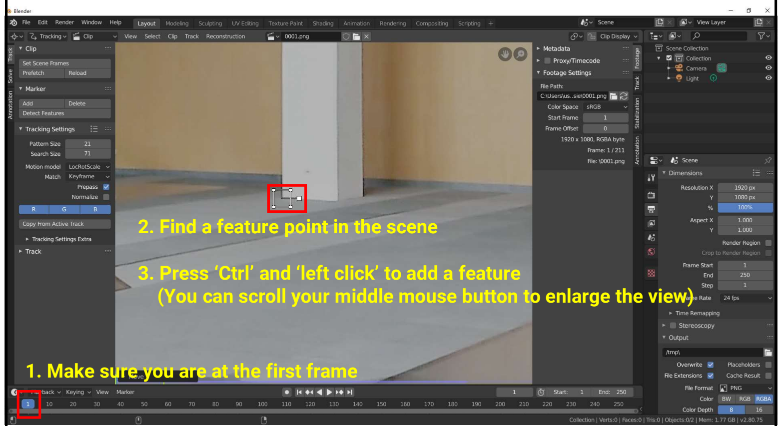
Set Input Data (2)



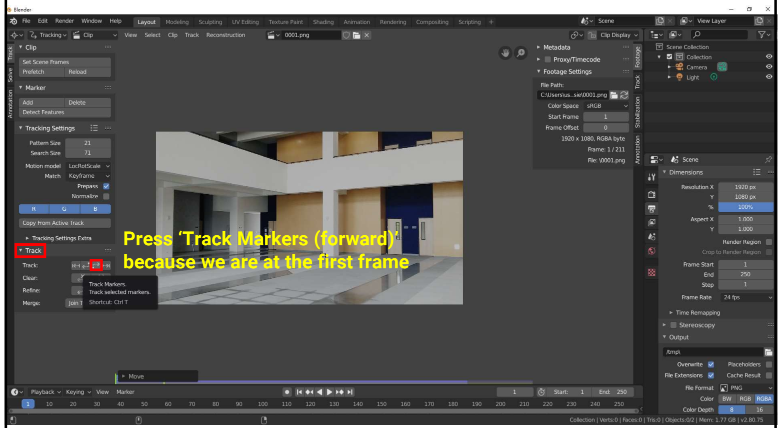
Feature Detection (1)



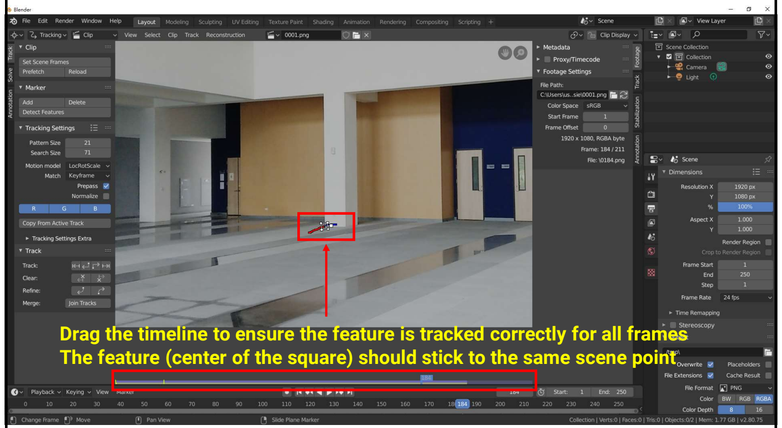
Feature Detection (2)



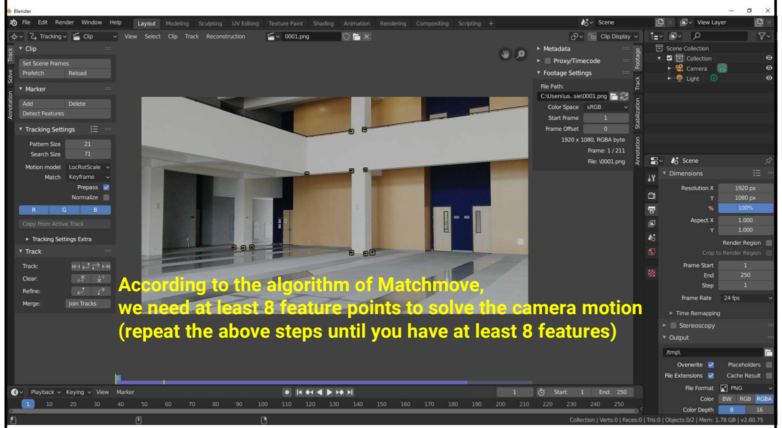
Feature Detection (3)



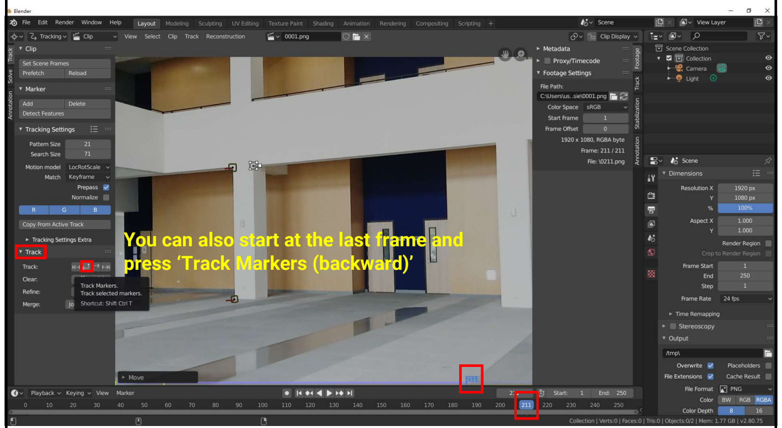
Feature Detection (4)



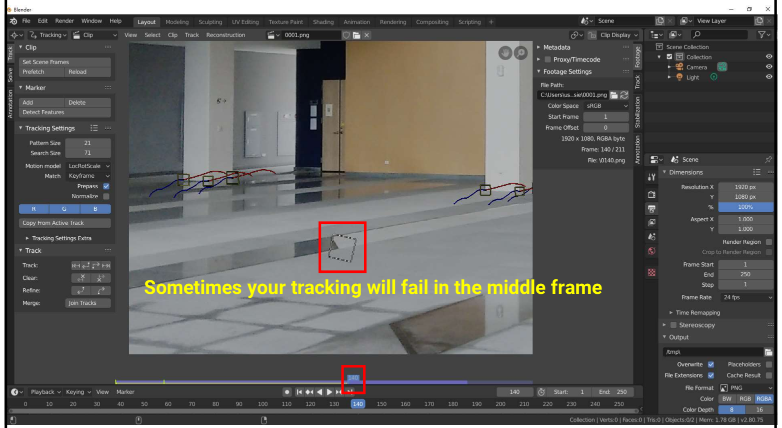
Feature Detection (5)



Feature Detection (6)

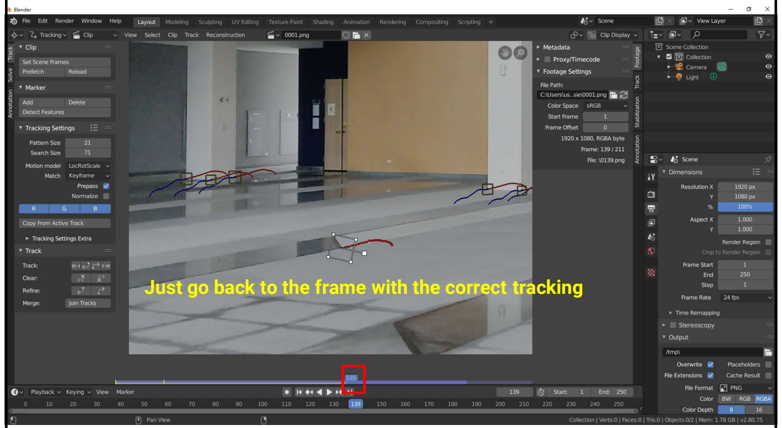


Feature Detection (7)



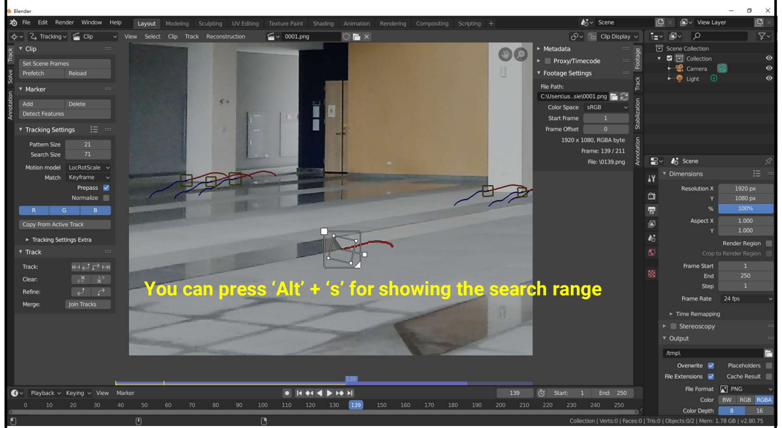
81

Feature Detection (8)



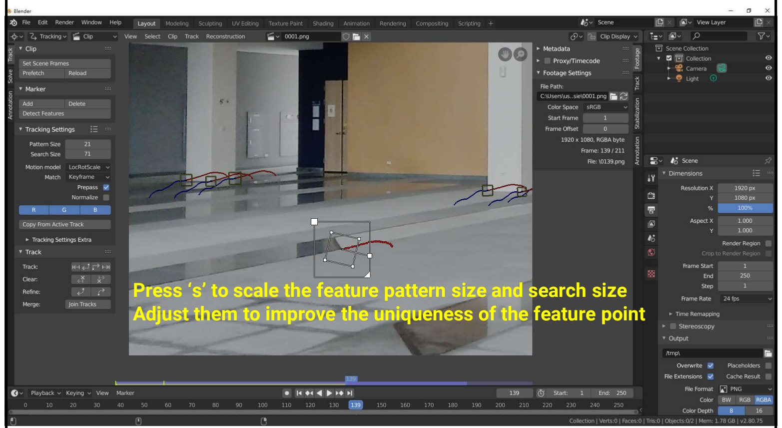
82

Feature Detection (9)



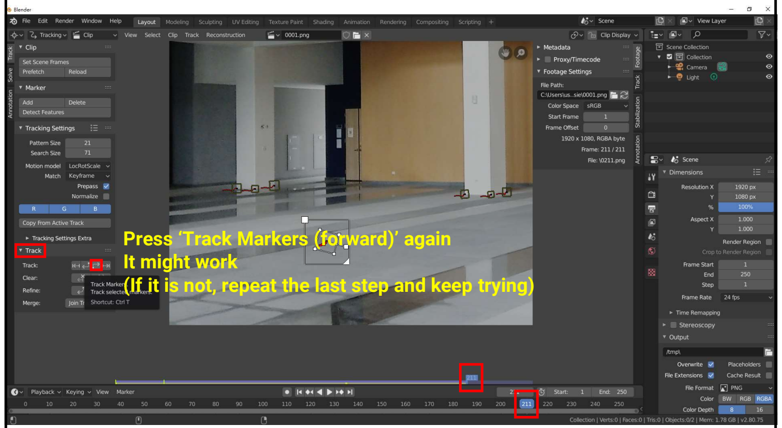
83

Feature Detection (10)



84

Feature Detection (11)

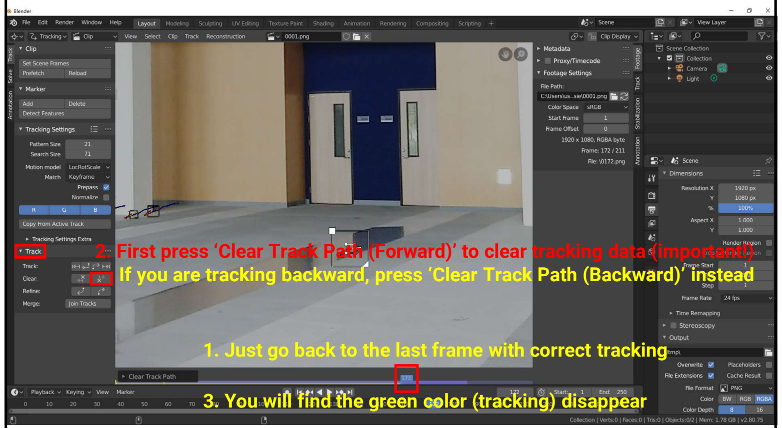


85

Feature Detection (12)

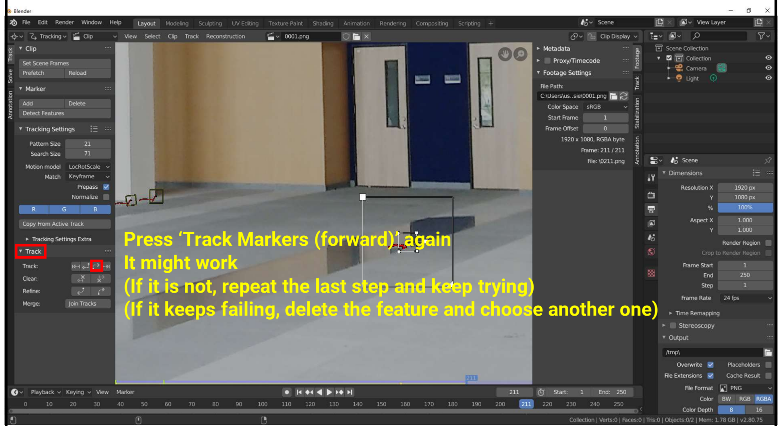
86

Feature Detection (13)



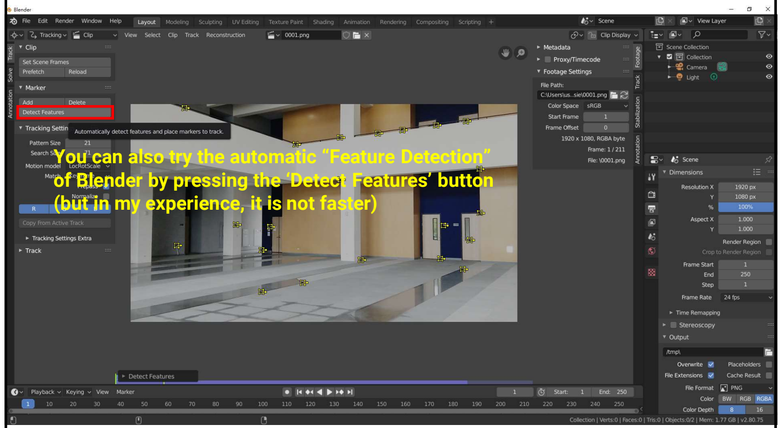
87

Feature Detection (14)



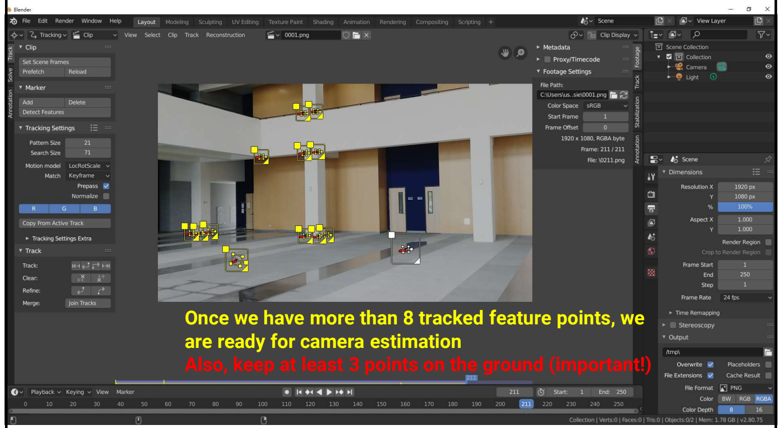
88

Feature Detection (15)



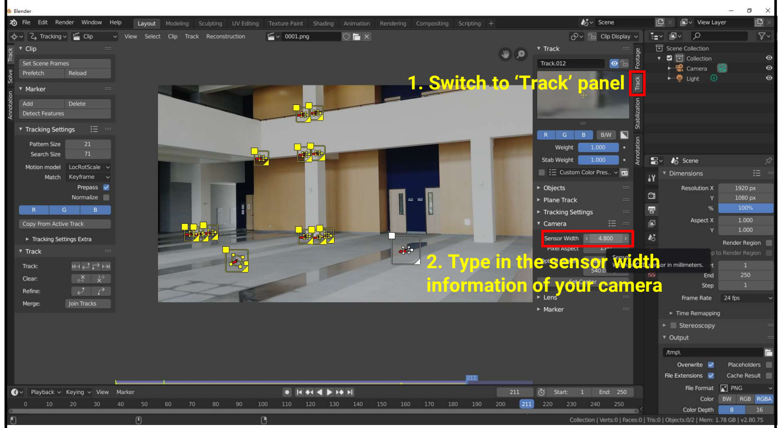
89

Feature Detection (16)



90

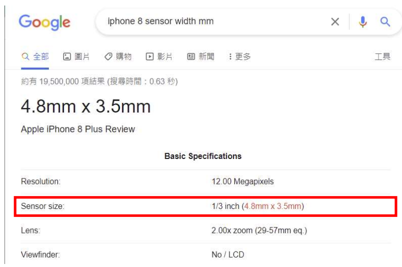
Setting Camera Parameters (1)



91

Camera Parameters

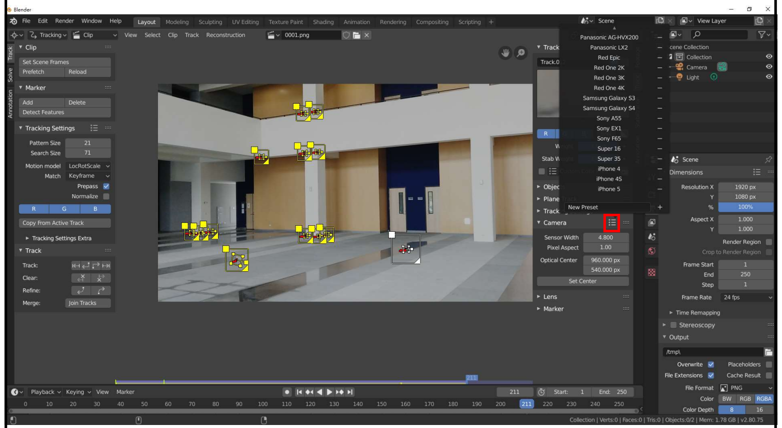
- **How do we know the camera parameters?**
  - Google it on the internet
    - https://www.photocounter.com.au/wp-content/uploads/2013/01/sensor-size-table.pdf
  - You can also use Blender's preset



Sensor "Type"	Imaging Area Dimensions			
	Diagonal (mm)	Width (mm)	Height (mm)	Area (mm^2)
1/6"	2.7	2.46	1.8	4.43
1/4"	4.5	3.6	2.7	9.72
1/3.6"	5.0	4.0	3.0	12.0
1/3.2"	5.68	4.54	3.42	15.53
1/3"	6.0	4.8	3.6	17.28
1/2.7"	6.72	5.37	4.04	21.69
1/2.5"	7.18	5.76	4.29	24.71
1/2.4"	7.66	5.92	4.57	27.05
1/2.33"	7.7	6.12	4.51	27.60
1/2.3"	7.8	6.17	4.55	28.07
1/2"	8.0	6.4	4.8	30.72
1/1.8"	8.93	7.18	5.32	38.20
1/1.75"	9.23	7.38	5.54	40.89
1/1.72"	9.25	7.40	5.55	41.07
1/1.7"	9.5	7.6	5.7	43.32
1/1.6"	10.07	8.08	6.01	48.56
2/3"	11.07	8.8	6.6	58.08
1"	16.0	12.8	9.6	122.88
4/3"	22.5	17.3	13.0	243.00

92

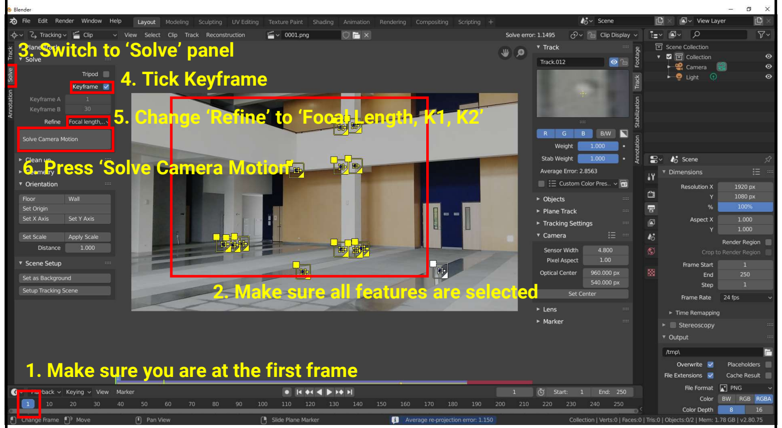
Setting Camera Parameters (2)



93

Multimedia Techniques and Applications 2022

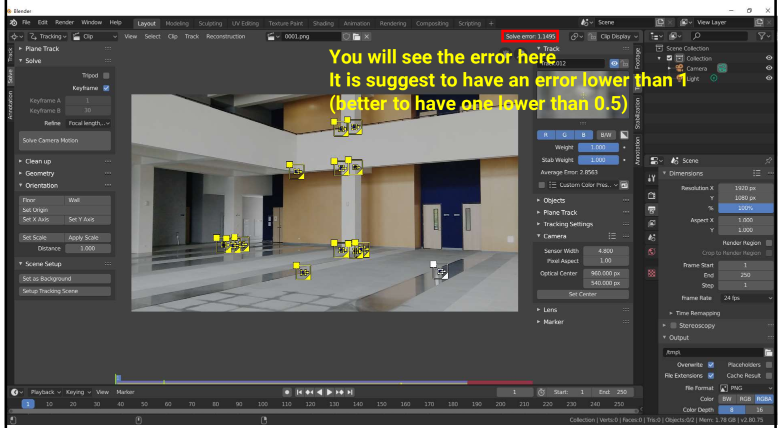
Solving Camera Motion (1)



94

Multimedia Techniques and Applications 2022

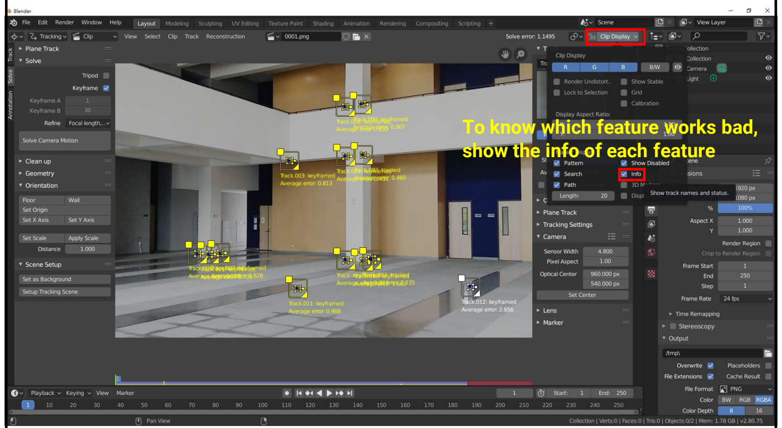
Solving Camera Motion (2)



95

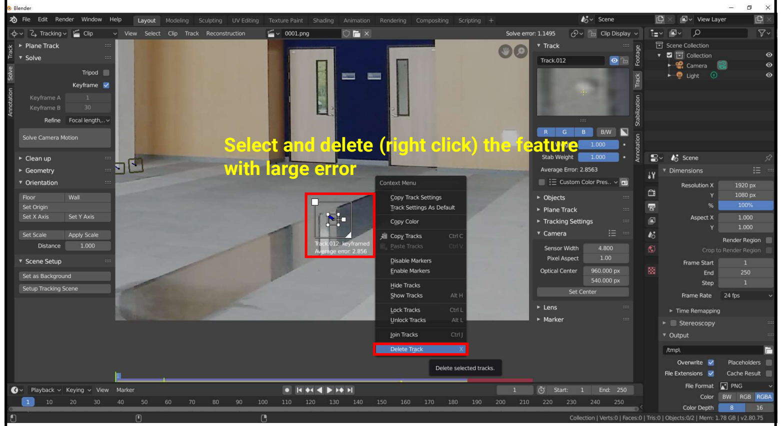
Multimedia Techniques and Applications 2022

Solving Camera Motion (3)



96

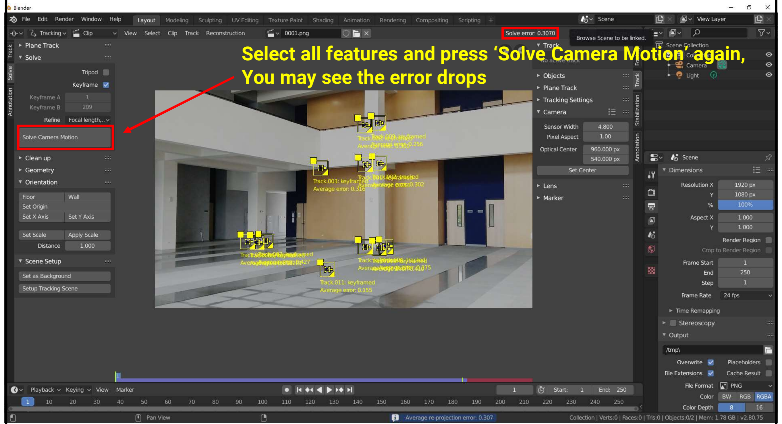
Solving Camera Motion (4)



97

Multimedia Techniques and Applications 2022

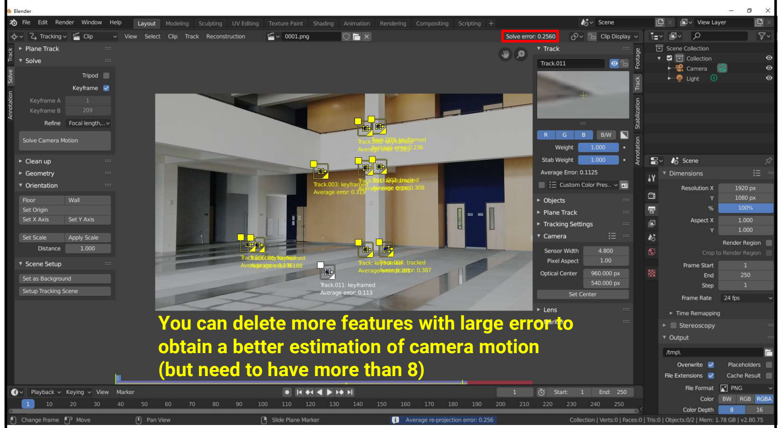
Solving Camera Motion (5)



98

Multimedia Techniques and Applications 2022

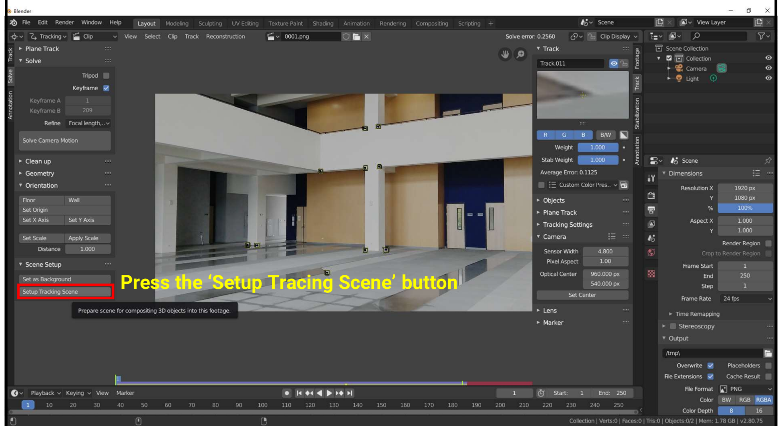
Solving Camera Motion (6)



99

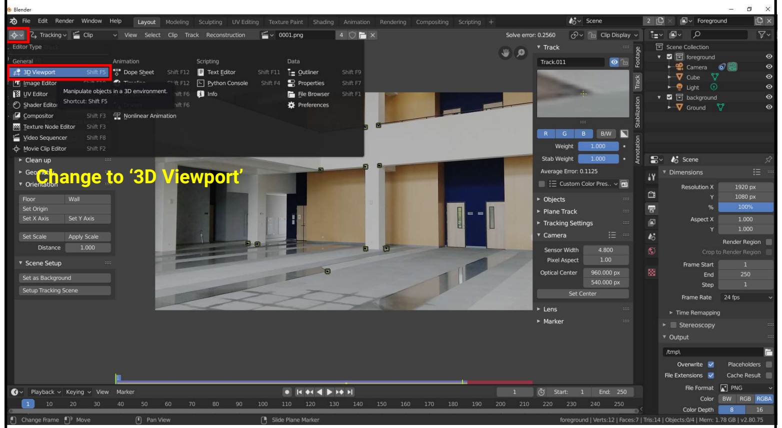
Multimedia Techniques and Applications 2022

Setup Tracking Scene (1)

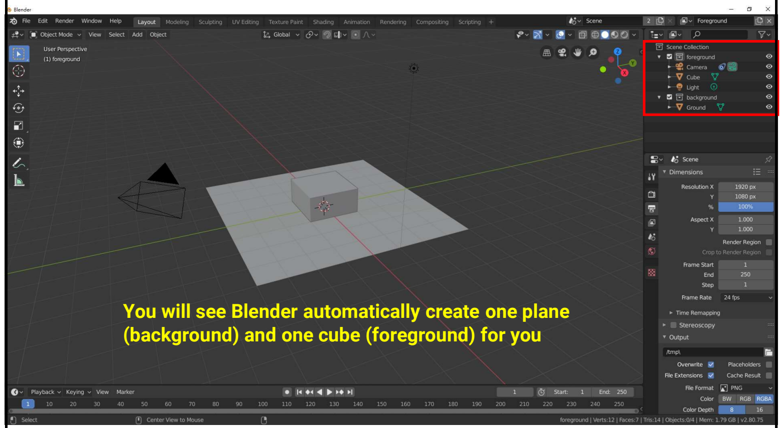


100

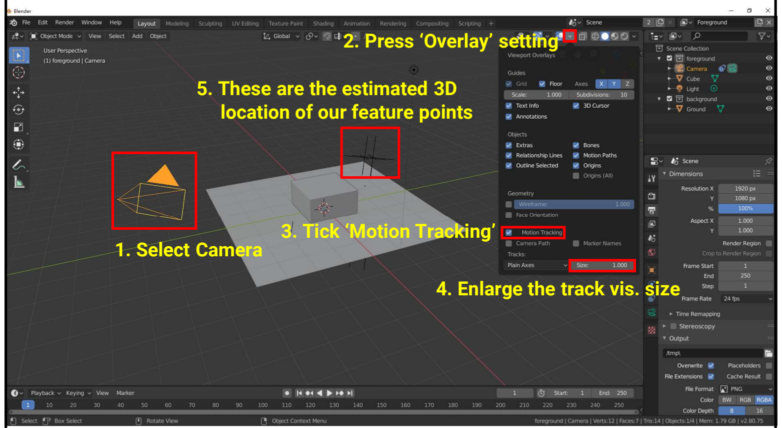
Setup Tracking Scene (2)



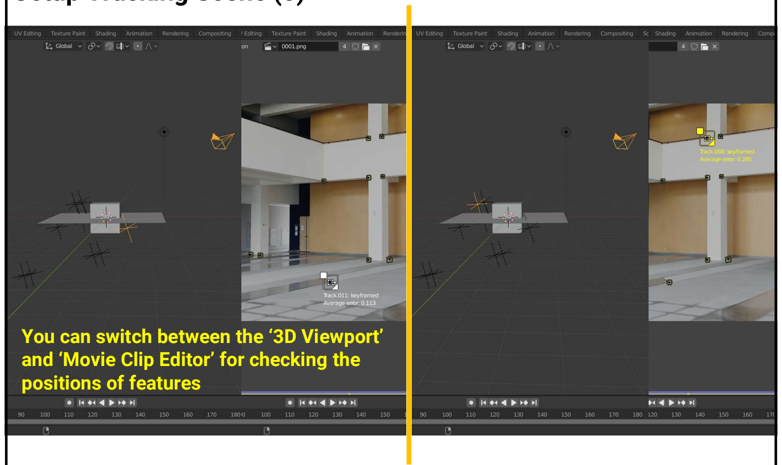
Setup Tracking Scene (3)



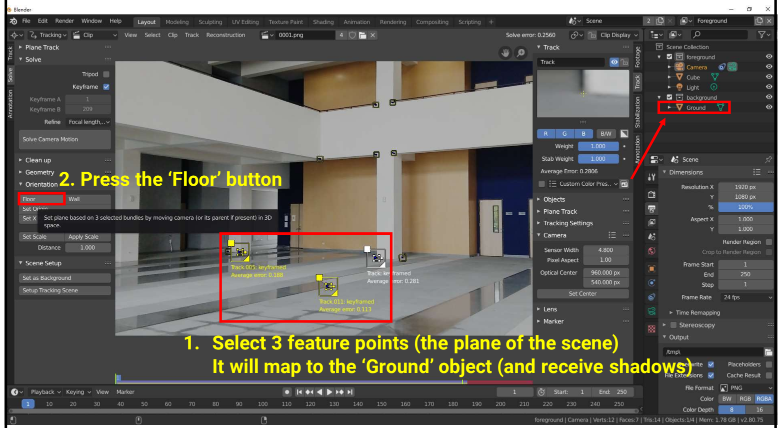
Setup Tracking Scene (4)



Setup Tracking Scene (5)

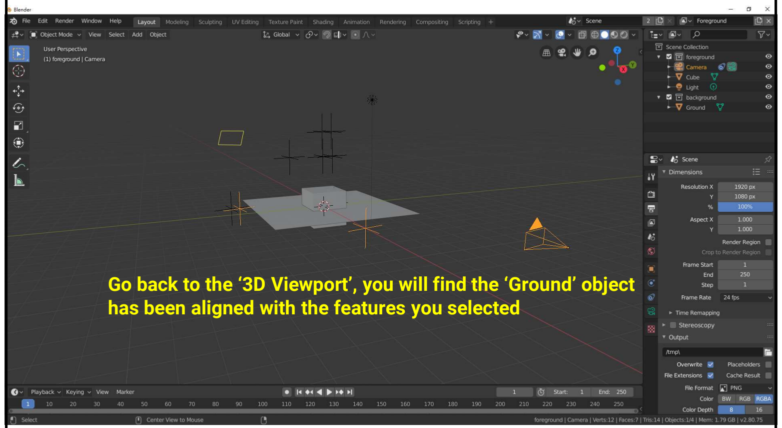


Setup Tracking Scene (6)



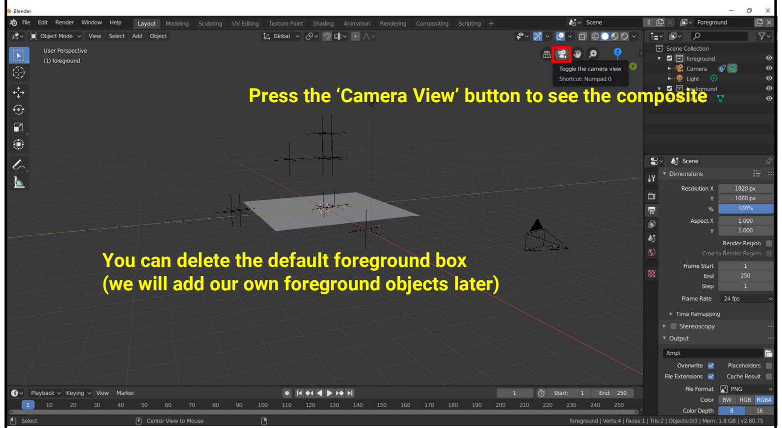
105

Setup Tracking Scene (7)



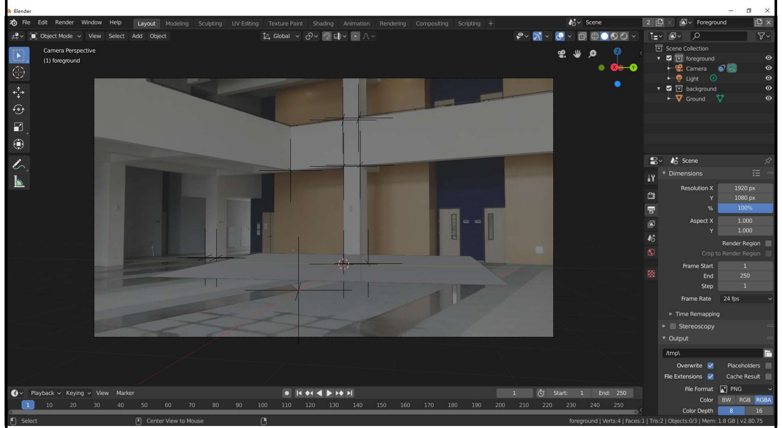
106

Setup Tracking Scene (8)



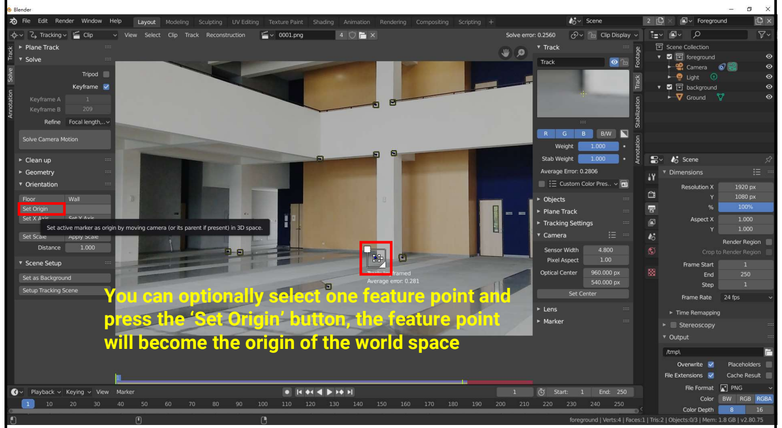
107

Setup Tracking Scene (9)

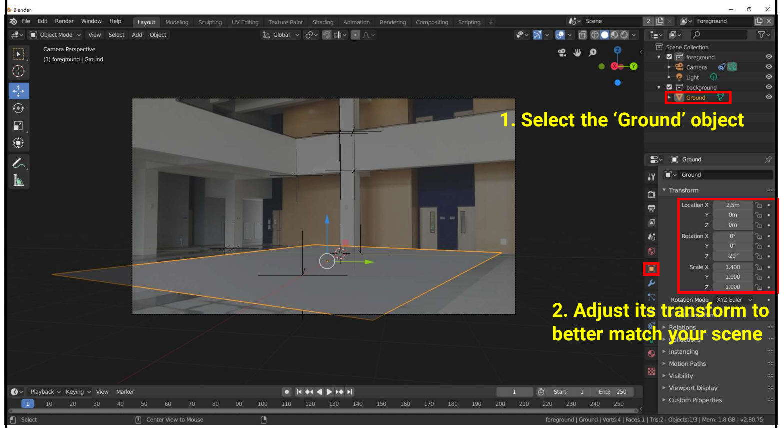


108

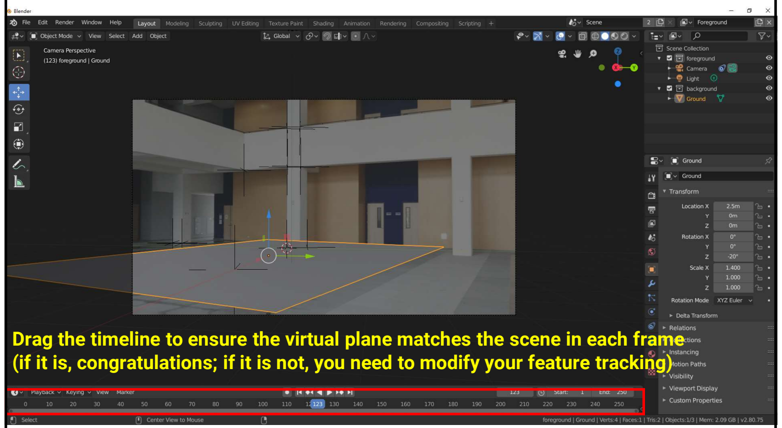
Setup Tracking Scene (10)



Setup Tracking Scene (11)



Setup Tracking Scene (12)

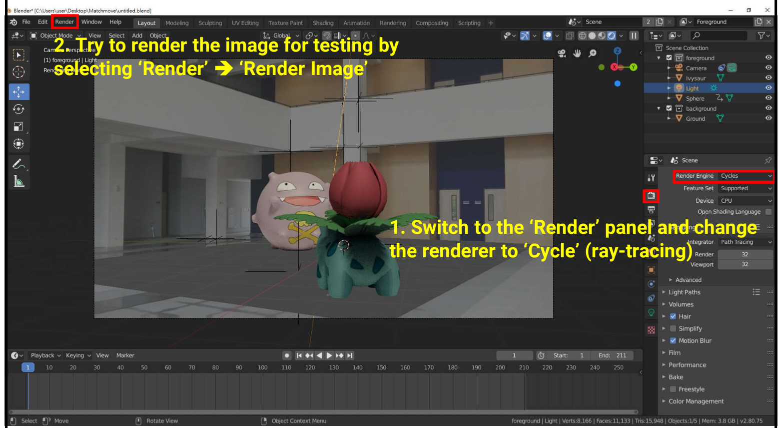


Add Virtual 3D Models (and Animations)

Add virtual 3D models and their animations

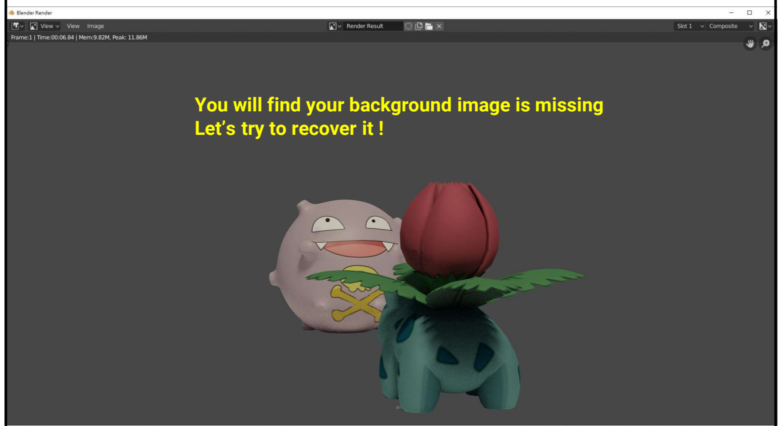
113

Try to Render the Frame (1)



114

Try to Render the Frame (2)

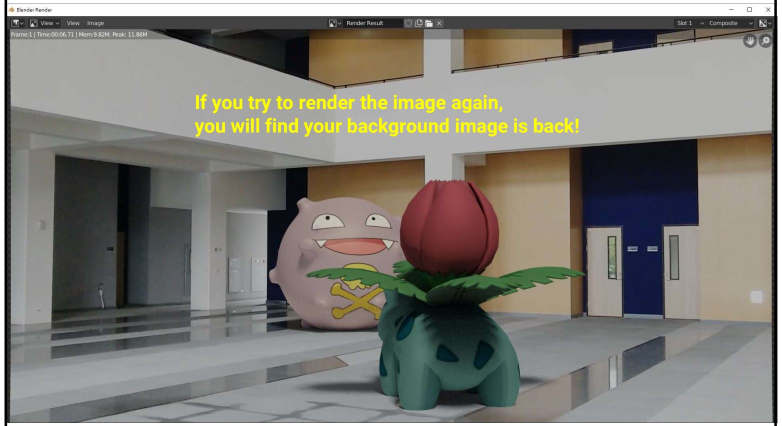


115

Try to Render the Frame (3)

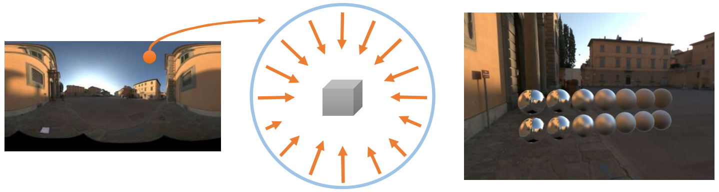
116

Try to Render the Frame (4)



Recap: Environment Lighting

- Environment light illuminates the scene from a **virtual sphere at infinite distance**
- The spherical energy distribution is usually represented with longitude-latitude images
- Also called **image-based lighting (IBL)**

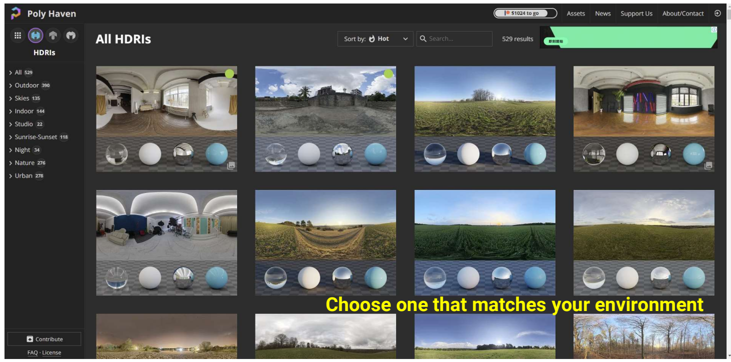


Recap: Environment Lighting

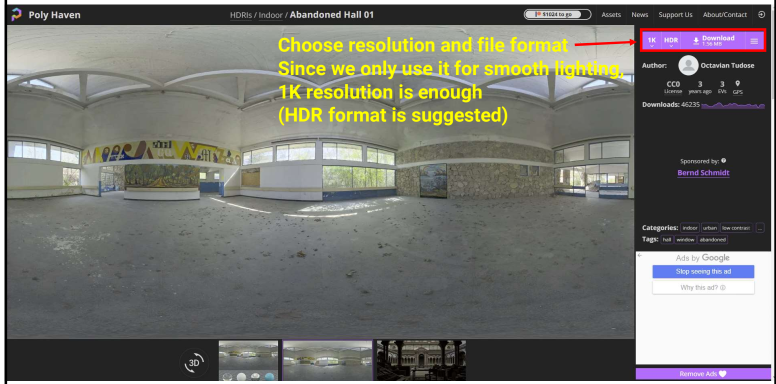
- Widely used in digital visual effects and film production

Environment Lighting Resource

- Download free HDR environment map on the internet
  https://polyhaven.com/hdris



Environment Lighting Resource (cont.)

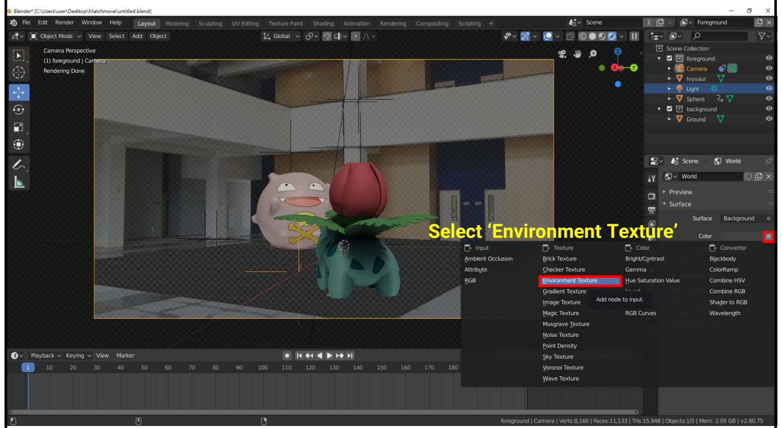


121

Add more realistic lighting (1)

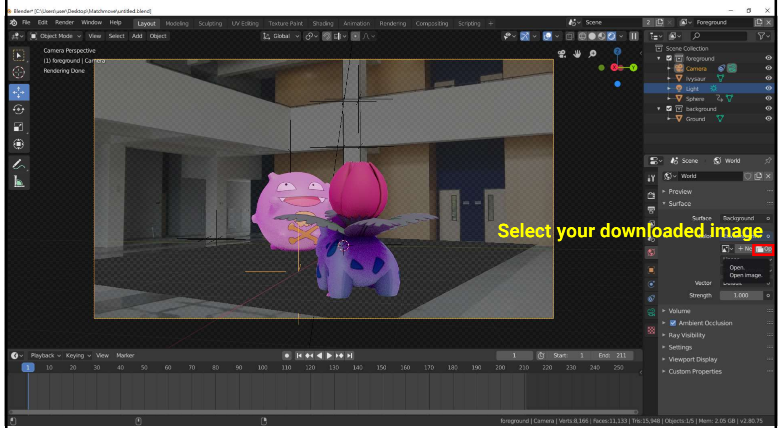
122

Add more realistic lighting (2)



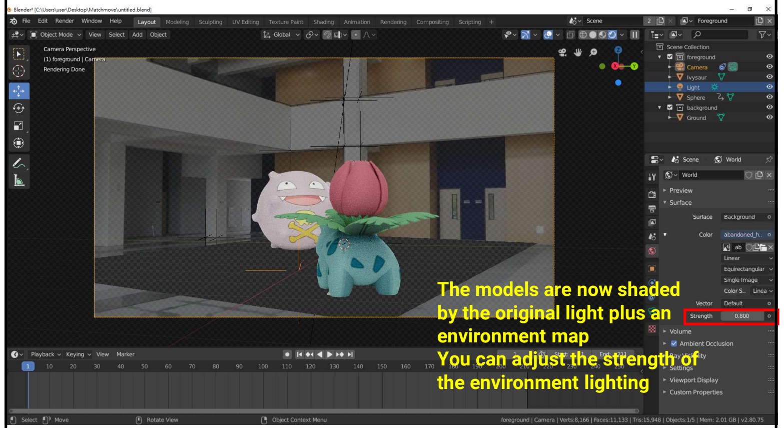
123

Add more realistic lighting (3)



124

Add more realistic lighting (4)

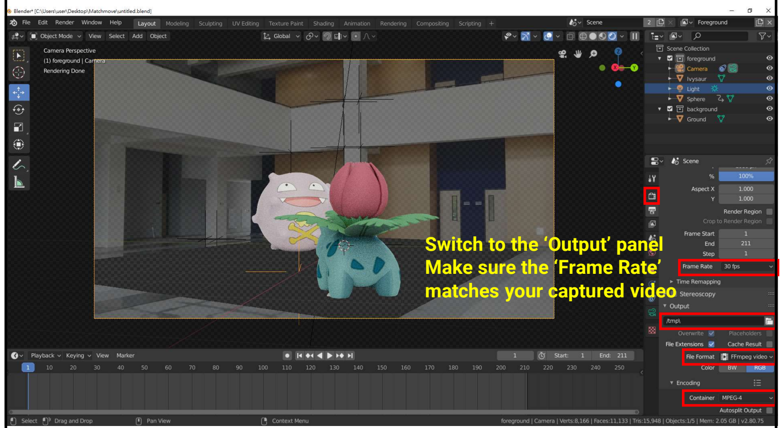


125

Output Composite Video

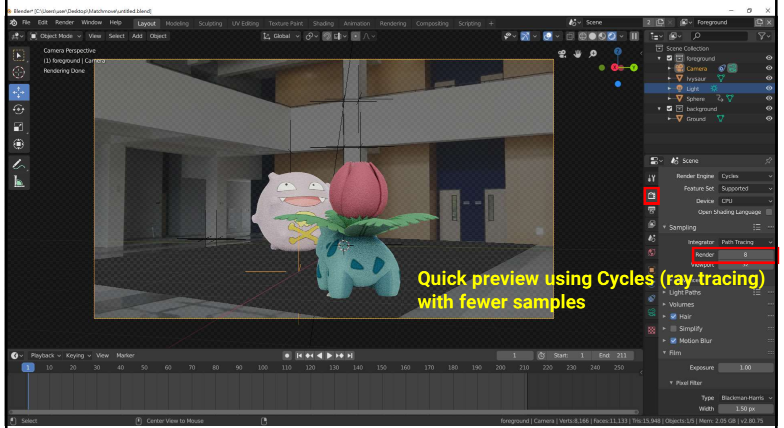
126

Set output configuration (1)



127

Set output configuration (2)

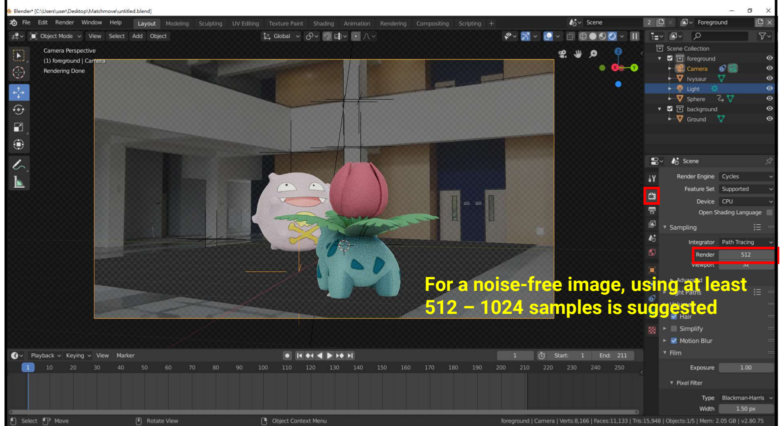


128

Render animation preview

Preview (check the poses and animations of the models)

Set to high-quality rendering



Final Output