

Supplementary Material for Non-Local Characterization of Scenery Images: Statistics, 3D Reasoning, and a Generative Model

Tamar Avraham and Michael Lindenbaum

Computer science department, Technion - I.I.T., Haifa 3200, Israel.
tammya,mic@cs.technion.ac.il

1 Introduction

This document is supplementary to the ECCV submission. Section 2 provides details about additional corrections and unifications of manual annotations using the Labelme toolbox [1]. Section 3 summarizes the occurrences of the different object labels annotated in the landscape images used in this work, divided into the two classes of foreground and background objects. In Section 4 more detailed calculations are provided for the 3D analysis discussed in Section 3 in the main paper.

2 Annotation Corrections

The Labelme toolbox [1] includes a dictionary that allows the user to collect annotations with similar semantic meaning under one name, e.g., the tag PERSON replaces a long list of free annotations like *a man*, *person walking*, *person crop*, *human figure tennisman*, *pedestrian*, *human*; The tag CAR replaces *red car*, *taxi*, *car front*, etc. This tagging is used also for correcting spelling mistakes, e.g., *bicycle* is replaced with BICYCLE.

When collecting the statistics presented in this work we used this dictionary, with a few small additions/changes, listed below:

- In the original dictionary there was one tag TREE replacing all words describing trees, both with single and plural meaning. We divided this group under two tags: TREE for words for single trees including, for instance, *palm tree*, *snowy tree*, *tree*, etc. and TREES for words describing plural trees such as *trees*, *palm trees*, *tree region*, etc.
- In the same manner we divided ROCK and ROCKS and PLANT and PLANTS.
- We recognized additional spelling mistakes and annotations that should have been grouped under existing tags: *volcano*, *rocky moutain*, *mountan*, and *moutain* were added to the tag MOUNTAIN. The phrase *rocky plain* was added under ROCKS. The words *brushes*, *brush*, *scrubland*, *ferns vegetation vineyard* and *undergrowth bruhses* were added under PLANTS. The terms

lake water and *pond* were added to the tag LAKE. The misspelled word *siky* was added to SKY. The word *paht* were added to tag PATH. The words *trres* and *forest* were added to TREES. The words *bluildings*, *buildins*, and *bouldings* were added to the tag BUILDING. The phrase *goose occluded* was added to tag GOOSE. The word *animals* was added to ANIMAL. The terms *skier*, *skier crop*, *biclist occluded*, *bicyclist*, and *bicyclists* were added to the tag PERSON. The word *machines* was added to MACHINE. The phrase *monolith crop* was added to MONOLITH. The word *ocean* was added to SEA. The word *sticks* was added to STICK. The word *van* was added to CAR. The term *fall branch* and *branch* were added to TREE (*tree branch* was already there). The phrase *gate occluded* was added to FENCE (that already included '*gate, fence*').

- We created a few new tags: the tag VALLEY including *valley*, *urban valley*, and *urban valey*, the tag PLAIN including *plain* and *urban plain*, the tag SHIP including *ship* and *ship occluded*, the tag WAVE including *wave*, *waves*, and *wave splash*, the tag FOGBANK including *fog banck* and *fog bank*, and the tag QUAY including *quay* and *jetty*.

3 Foreground and Background labels in Labelme landscape images

Labelme images, categorized under *coast*, *mountain*, and *open country* [2] were used for collecting statistics about scenery images. We divided the objects annotated in those images to background objects and foreground images. Fig. 1 summarizes the occurrences of each label.

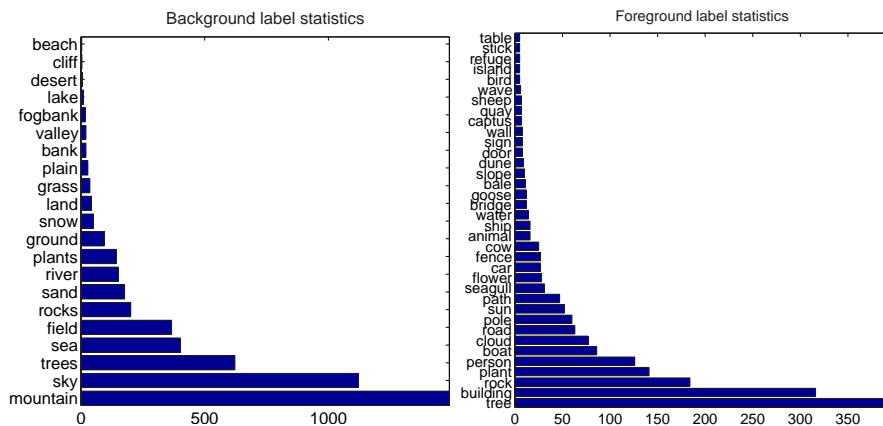


Fig. 1. Objects are semantically divided into *background objects* and *foreground objects*. Left: the background labels and occurrences of each in the Labelme landscape images; right: the foreground labels and occurrences of each (including all identities with at least 5 occurrences)

4 From an angle on a plane to its projection on the image plane

4.1 From an angle on a flat terrain to its projection on the image plane

A pinhole camera is located at height h from the ground. Consider two coordinate systems as depicted in Fig. 2. The first is a 3D coordinate system (X_1, X_2, X_3) , with an origin at the pinhole O . The second is a 2D coordinate system (Y_1, Y_2) whose origin is in the center of the image plane. Consider a short line PQ on a flat terrain, the angle of which is θ relative to the axis X_1 . In the 3D coordinate system,

$$\begin{aligned} P(x, -h, z) \quad \text{and} \\ Q(x + \Delta \cos \theta, -h, z + \Delta \sin \theta) . \end{aligned}$$

In general, a point (x_1, x_2, x_3) in the 3D coordinate system (X_1, X_2, X_3) is mapped to the point

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \frac{f}{x_3} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \quad (1)$$

in the image plane (Y_1, Y_2) (perspective projection), where f is the camera focal length [3]. Therefore, the projection of points P and Q on the image plane are in

$$\begin{aligned} P' \left(\frac{fx}{z}, -\frac{fh}{z} \right) \quad \text{and} \\ Q' \left(\frac{f(x + \Delta \cos \theta)}{z + \Delta \sin \theta}, -\frac{fh}{z + \Delta \sin \theta} \right) , \end{aligned}$$

and

$$\tan \theta' = \frac{Y_2(Q') - Y_2(P')}{Y_1(Q') - Y_1(P')} = \frac{h \tan \theta}{z - x \tan \theta} . \quad (2)$$

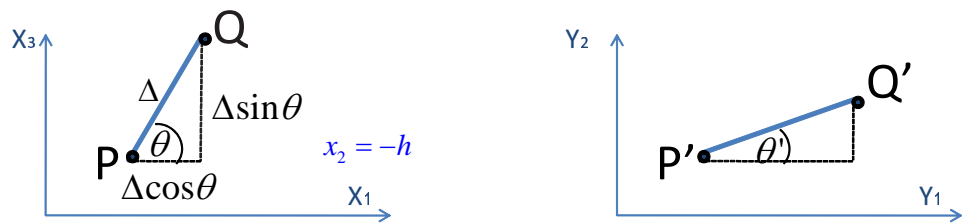
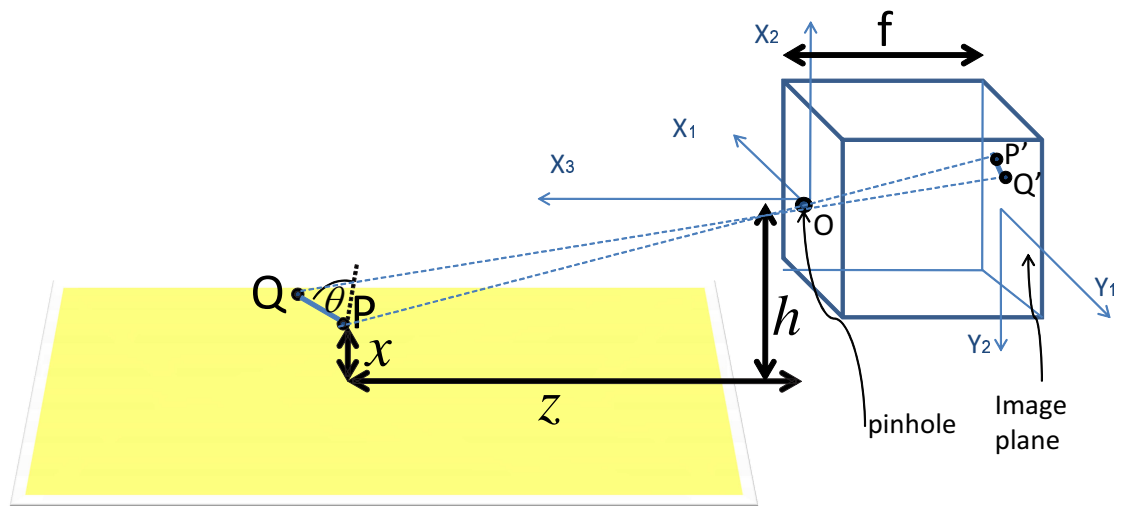


Fig. 2. The image of a line segment on the ground according to a pinhole camera model. PQ is a line on a flat surface, filmed by a pinhole camera, the center of which is of height h from the ground. The (X_1, X_2, X_3) coordinate system is centered in the pinhole O . A 2D coordinate system (Y_1, Y_2) is centered in the image plane. PQ is oriented in an angle θ relative to the X_1 axis. $P'Q'$ is the projection of PQ on the image plane. See text for more details and for the relation between θ' , the angle between $P'Q'$ and the Y_1 axis, and θ

4.2 From an angle on a rotated slope to its projection on the image plane

Now, consider the plane discussed in the previous section being tilted by a slope angle φ , and rotated at angle ω . See Fig. 3. Let the coordinate of the point P in the 3D coordinate system (X_1, X_2, X_3) centered in O , the pinhole, be $P(x, H, z)$. Then Q is ¹

$$Q \left(x + \Delta(\cos \theta \cos \omega - \sin \theta \cos \varphi \sin \omega), \right. \\ \left. H + \Delta \sin \theta \sin \varphi, \right. \\ \left. z + \Delta(\cos \theta \sin \omega + \sin \theta \cos \varphi \cos \omega) \right) .$$

The projections of P and Q in the image plane (Y_1, Y_2) are then

$$P' \left(f \frac{x}{z}, f \frac{H}{z} \right) \text{ and} \\ Q' \left(f \frac{x + \Delta(\cos \theta \cos \omega - \sin \theta \cos \varphi \sin \omega)}{z + \Delta(\cos \theta \sin \omega + \sin \theta \cos \varphi \cos \omega)}, f \frac{H + \Delta \sin \theta \sin \varphi}{z + \Delta(\cos \theta \sin \omega + \sin \theta \cos \varphi \cos \omega)} \right) ,$$

and

$$\tan \theta' = \frac{Y_2(Q') - Y_2(P')}{Y_1(Q') - Y_1(P')} \quad (3) \\ = \frac{H(\cos \theta \sin \omega + \sin \theta \cos \varphi \cos \omega) - z \sin \theta \sin \varphi}{x(\cos \theta \sin \omega + \sin \theta \cos \varphi \cos \omega) - z(\cos \theta \cos \omega - \sin \theta \cos \varphi \sin \omega)} .$$

Note that the case discussed in the previous section is the private case of $\varphi = 0$ and $\omega = 0$.

¹ Consider P to be in $(0, 0, 0)$, and Q in $Q(\Delta \cos \theta, 0, \Delta \sin \theta)$. After tilting the plane, P stays in $(0, 0, 0)$, and Q moves to $Q(\Delta \cos \theta, \Delta \sin \theta \sin \varphi, \Delta \sin \theta \cos \varphi)$. Rotating in the $X_1 - X_3$ plane around P moves Q to $Q(\Delta(\cos \theta \cos \omega - \sin \theta \cos \varphi \sin \omega), \Delta \sin \theta \sin \varphi, \Delta(\cos \theta \sin \omega + \sin \theta \cos \varphi \cos \omega))$. Finally, the entire plane is translated bringing P to $P(x, H, z)$ and Q to $Q(x + \Delta(\cos \theta \cos \omega - \sin \theta \cos \varphi \sin \omega), H + \Delta \sin \theta \sin \varphi, z + \Delta(\cos \theta \sin \omega + \sin \theta \cos \varphi \cos \omega))$.

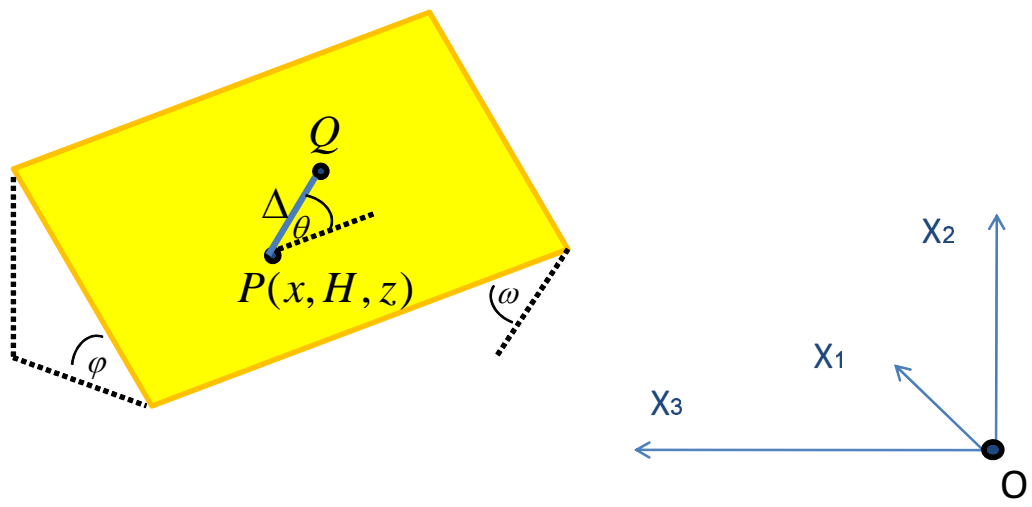


Fig. 3. The line segment PQ now lies on a flat tilted and rotated surface. The tilt of the surface forms a slope angle φ with the $X_1 - X_3$ plane. The surface is rotated forming an angle ω with the X_1 axis

References

1. Russell, B., Torralba, A.: Labelme: a database and web-based tool for image annotation. *IJCV* **77** (2008) 157–173
2. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. *IJCV* **42** (2001) 145–175
3. Forsyth, D.A., Ponce, J.: *Computer Vision, A Modern Approach*. Prentice Hall (2003)