

Wall, Floor, Ceiling, Object Region Identification from Single Image

Zhong-Ju Zhang
Stanford University
junez@stanford.edu

Abstract

It is often helpful to identify large, relevant regions in an image; this can help to facilitate applications such as object recognition, object tracking, and 3D reconstruction. We explored a simple algorithm of using K-means to cluster images of indoor scenes into 4 regions roughly corresponding to 1) ceiling, 2) wall, 3) floor, and 4) objects (non-ceiling, wall, floor) regions; then we used logistic regression to identify the segmented clusters to the corresponding regions. K-means segmentation performs well for uncluttered scene with large and uninterrupted wall and floor areas. Ceiling and object regions are harder to segment. Identification of segmented region is poor, achieving only 50% accuracy using logistic regression.

Keywords: image segmentation, region identification, K-means

1 Introduction

In image processing, it is often useful to be able to separate and classify large regions in an image. Identifying image regions as ground or sky was useful for Hoiem et al. in learning 3D geometric context from a single image [1]. Identifying large regions can also improve image tracking by allowing the algorithm to focus more effectively on objects of interest.

Hoiem et al. used supervised learning algorithm with 78D feature vector to segment and identify regions in outdoor scenes. In this paper, we choose to analyze indoor scenes in order to identify regions corresponding to 1) ceiling, 2) wall, 3) floor, and 4) objects (non-ceiling, wall, floor regions) by employing fast K-means clustering algorithm for segmentation and logistics regression for region identification.

2 Segmentation

2.1 K-Means Algorithm

K-means clustering algorithm does not make any probabilistic assumptions on the data. One disadvantage of K-means is that users have to guess the total number of clusters, k , in the image before segmentation. Its performance is not as good as more sophisticated clustering algorithms such as mean shift [2]. However, for our purpose of identifying large regions (e.g. wall, floor), we argue that a sophisticated clustering algorithm is unnecessary as individual objects in the scene all belong to one region—object.

We applied K-means clustering ($K=4$) over feature space to find the ceiling, wall, floor, and object region in indoor scenes.

2.2 Features

We tried to choose a feature space that is descriptive of the image by incorporating colour, spatial, and texture descriptors.

2.2.1 Texture Energy

We used Law's Texture Energy algorithm to compute the texture energy of each pixel. The 1D texture masks are:

$$L5(level) = [1 \quad 4 \quad 6 \quad 4 \quad 1]$$

$$E5(edge) = [-1 \quad -2 \quad 0 \quad -2 \quad -1]$$

$$S5(spot) = [-1 \quad 0 \quad 2 \quad 0 \quad -1]$$

$$R5(ripple) = [1 \quad -4 \quad 6 \quad -4 \quad 1]$$

Nine 5x5 2D texture masks can then be constructed by combining the 1D masks: L5E5, L5R5, E5S5, S5S5, R5R5, L5S5, E5E5, E5R5, S5R5 [3].

We convolved the blue intensity channel of the image with the 9 different 5x5 Law's texture matrix. The texture energy of a pixel is the sum of the absolute value of the convolution result of its 8 nearest neighbours.

Experimentally, incorporating the L5E5 and L5R5 Law's texture energy gave the best segmentation result as shown in Figure 1, Figure 2, Figure 3.



Figure 1. a) Test image 1, b) K-means clustering



Figure 2. a) Test image 2, b) K-means clustering



Figure 3. a) Test image 3, b) K-means clustering

Table 1. Features for each pixel

Feature Descriptors	Num
Colours RGB values (scaled)	3
Spatial X, Y values (scaled)	2
Texture L5E5, L5R5 (scaled)	2

The colour, spatial, and texture values are scaled by weights so that R, G, B, X, Y, and the L5E5 texture value are approximately on the same scale. The L5R5 value is scaled to be 1/6 the L5E5 value.

3 Region Labelling

Major regions in the image can be found using segmentation. However, clustering assignment is meaningless; with every run, different clusters are found in difference sequences. For instance, the 3rd cluster corresponds to floor pixel, while in another clustering iteration, the 3rd cluster may correspond to wall pixels. We tried using logistic regression to identify the segmented clusters as 1) ceiling, 2) wall, 3) floor, and 4) objects (non-ceiling, wall, floor regions) by solving for parameters that maximize the likelihood a given set of features belongs to one of the 4 major regions. We choose a discriminative learning algorithm as opposed to a generative algorithm (e.g. Gaussian Discriminant Analysis) because we did not want to make any probabilistic assumption on the distribution of the feature descriptors.

3.1 Training using pre-labelled image

To test the performance of logistic regression for identification, we hand labelled the test image, shown

in Figure 4, to use as region labels for training test image 1.



Figure 4. Hand labelled regions for test image 1

We created 4 training labels, $y_{ceiling}, y_{floor}, y_{wall}, y_{object}$; the label is 1 if the pixel belongs to that region, 0 if it does not. We use the following notation for describing a feature point: $x_j^i \in \mathcal{R}^7, i = 1 \dots mn, j = 1..4$, i corresponds to the total number of pixels in an image that is $m \times n$, j corresponds to the 4 segmented clusters. Using logistic regression, we find 4 sets of parameters:

$$\theta_{ceiling} := \arg \max_{\theta} \log \sum_{i=1}^{mn} P(y_{ceiling}^i | x^i; \theta_{ceiling})$$

$$\theta_{floor} := \arg \max_{\theta} \log \sum_{i=1}^{mn} P(y_{floor}^i | x^i; \theta_{floor})$$

$$\theta_{wall} := \arg \max_{\theta} \log \sum_{i=1}^{mn} P(y_{wall}^i | x^i; \theta_{wall})$$

$$\theta_{object} := \arg \max_{\theta} \log \sum_{i=1}^{mn} P(y_{object}^i | x^i; \theta_{object})$$

3.1.1 Labelling

After obtaining the parameters, for each region, we then compute:

$$p(y_{wall} | x_{cluster1}^{centroid}; \theta_{wall})$$

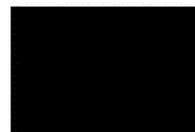
$$p(y_{wall} | x_{cluster2}^{centroid}; \theta_{wall})$$

$$p(y_{wall} | x_{cluster3}^{centroid}; \theta_{wall})$$

$$p(y_{wall} | x_{cluster4}^{centroid}; \theta_{wall})$$

where $x_{cluster1}^{centroid}$ is the centroid point identified by K-means for cluster 1. We do this separately for all four regions: ceiling, wall, floor, and objects. We assigned the region label to the cluster with the highest probability. A secondary condition is that the probability need to be larger than 0.25 to avoid false positives. Figure 5 shows the result of the identification for test image 1.

This is class 1=floor with probability0%



This is class 2=ceiling with probability0%



This is class 3=wall with probability99%



This is class 4=object with probability56%



Figure 5. Region labelling of test image 1. White region are clusters that is assigned the corresponding class label.

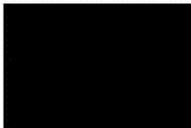
Even for a trivial case, logistic regression result is quiet poor, despite the fact that the segmentation results were quiet good for the major regions of the image (see Figure 1).

Using the parameters obtained by using the label shown in Figure 4, we then tried to cluster the different but similar looking test image 3 (see Figure 3)

This is class 1=floor with probability0%



This is class 2=ceiling with probability0%



This is class 3=wall with probability85%



This is class 4=object with probability59%



Figure 6. Region labelling of test image 3

3.2 Training using structure labelled image

Instead of using pre-labelled image to train the identifier, we used an artificial region label, shown in Figure 7, based on the assumption that the ceiling is above the wall, and the floor is below the wall in the image.

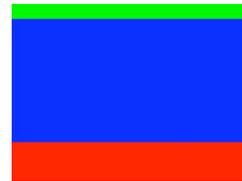
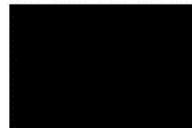


Figure 7. Assumed structure label.

Since the assumed region label does not contain any information regarding object region, we can only find parameters corresponding to ceiling, floor, and wall using logistic regression.

For identification, we assign label ceiling, floor, and wall to clusters that have the maximum probabilities of being these regions. For identifying the object cluster, we choose the cluster that has the minimum probability of being ceiling, floor, and wall region. Figure 8,9 shows the result of identification using the assumed structure region label.

This is class 1=floor with probability0%



This is class 2=ceiling with probability0%



This is class 3=wall with probability100%



This is class 4=object with probability0%



Figure 8. Region labelling of test image 1 using structure labelled image

This is class 1=floor with probability0%



This is class 2=ceiling with probability0%



This is class 3=wall with probability67%



This is class 4=object with probability0%



Figure 9. Region labelling of test image 2 using structure labelled image

4 Result

K-means segmentation performs well for a well-behaved image with large wall and floor area. For all the test images we tried, the ceiling was clustered together with the wall because it has the same texture and colour as the wall. Without geometric cues, it is unlikely that any clustering algorithm will be able to separately cluster ceiling from wall region. The object regions are clustered with varying success. Depending on the amount of objects in a space, they could be clustered as one or more clusters.

The region labelling results are quiet poor even for images with good segmentation results. Ceiling and floor region are usually never identified because many probability values end up being zero. Clusters are usually identified correctly as wall regions with high probability. Object region identification success can vary. The accuracy of object region identification is dependent on how well K-means segmentation separated objects from floor, wall clusters. In most cases, object regions can also partially contain wall and/or floor regions as well.

5 Future Work

It is apparent that conventional logistic regression is not a good identifier. It would be useful to try generative learning algorithms (e.g. naïve bayes) for identification. A more descriptive feature vector can also help.

6 References

- [1] D. Hoiem, A. Efros, M. Herbert, "Geometric Context from a Single Image", www.cs.cmu.edu/~dhoiem/publications/Hoiem_Geometric.pdf
- [2] <http://www.caip.rutgers.edu/riul/research/code.html>
- [3] http://www.cse.msu.edu/~stockman/CV/F05Lectures/Weeks1_9/week06-texture-LS.ppt