

Unsupervised Segmentation of Color-Texture Regions in Image and Video

Yining Deng, B.S. Manjunath

Presented by
Chen-hsiu Huang

January 17, 2002

Abstract

In this paper, a new method for unsupervised segmentation of color-texture regions in images and video is presented. This method, which referred to as JSEG, consists of two independent steps:

1. Color Quantization

Colors in the image are quantized to several representative classes that can be used to differentiate regions in the image.

2. Spatial Segmentation

A region growing method is then used to segment the image and a similar approach is applied to video sequences. Besides, an additional region tracking scheme is embedded into the region growing process.

Image Homogeneity

In order to identify image homogeneity, the following assumptions about the image are made:

1. Each image contains a set of approximately homogeneous color-texture regions.
2. The color information in each image region can be represented by a set of few quantized colors.
3. The colors between two neighboring regions are distinguishable – basic assumption in any color image segmentation algorithm.

Segmentation Example

Online demo can be found at <http://maya.ece.ucsb.edu/JSEG/>



Original 352x240 MPEG-1 video
frame



Segmentation with $P_{quant} = 100$,
 $P_{scal} = automatic$, $P_{merge} = 0.4$

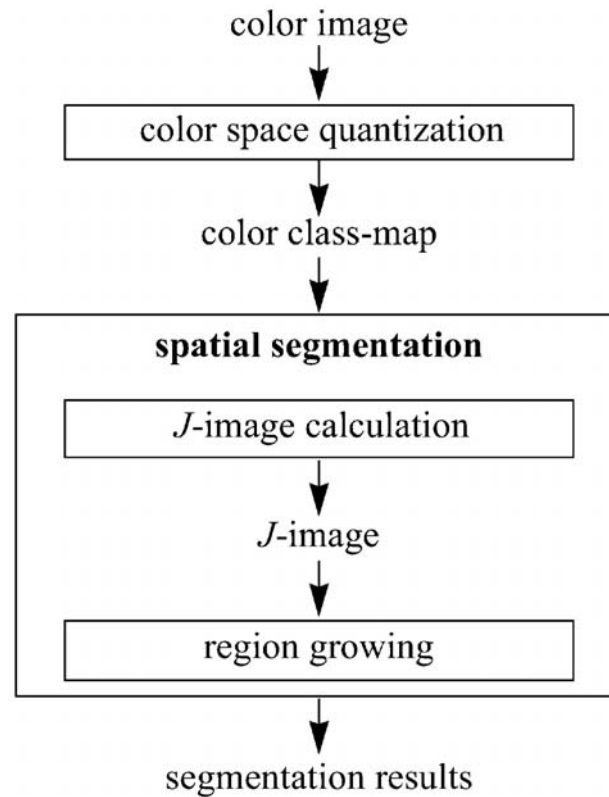
Color Quantization

- First, colors in the image are quantized to several representative classes that can be used to differentiate regions in the image. This is performed in the color space without considering the spatial distributions of the colors.
- Then, the image pixel values are replaced by their corresponding color class labels, thus forming a class-map of the image.
- In the second stage, spatial segmentation is performed directly on this class-map without considering the corresponding pixel color similarity.

Spatial Segmentation

- Introduce a new criterion for image segmentation. This criterion involves minimizing a cost associated with the partitioning of the image based on pixel labels.
- A practical algorithm is suggested toward achieving this segmentation objective. The notion of "*J*-images" is introduced. *J*-images correspond to measurements of local homogeneities at different scales, which can indicate potential boundary locations.
- A spatial segmentation algorithm is then described, which grows regions from seed areas of the *J*-images to achieve the final segmentation.

Schematic of the JSEG algorithm



Color-maps After Quantization

Following quantization, the quantized colors are as-signed labels. A color class is the set of image pixels quantized to the same color. The image pixel colors are replaced by their corresponding color class labels. The newly constructed image of labels is called a class-map.

Examples of a class-map are shown:

```

+++++0000
+++++0000
+++++0000
+++++0000
+++++0000
+++++*****
+++++*****
+++++*****
+++++*****

```

class-map 1
 $J=1.720$

```

+*+*+*+*+
0+0+0+0+0
+*+*+*+*+
0+0+0+0+0
+*+*+*+*+
0+0+0+0+0
+*+*+*+*+
0+0+0+0+0
+*+*+*+*+

```

class-map 2
 $J=0$

```

+++++*0*0
+++++0*0*
+++++*0*0
+++++0*0*
+++++*0*0
+++++*0*0
+++++0*0*
+++++*0*0
+++++0*0*

```

class-map 3
 $J=0.855$

Criterion for "Good" Segmentation (1/2)

- Let Z be the set of all N data points in a class-map. Let $Z = (z, y), z \in Z$, and m be the mean,

$$m = \frac{1}{N} \sum_{z \in Z} z$$

- Suppose Z is classified into C classes, $Z_i, i = 1, \dots, C$. Let m_i be the mean of the N_i data points of class Z_i ,

$$m_i = \frac{1}{N_i} \sum_{z \in Z_i} z$$

- Let

$$S_T = \sum_{z \in Z} \|z - m\|^2$$

and

$$S_W = \sum_{i=1}^C S_i = \sum_{i=1}^C \sum_{z \in Z_i} \|z - m_i\|^2$$

- S_W is the total variance of points belonging to the same class.

Define:

$$J = \frac{S_T - S_W}{S_W}$$

Criterion for "GOOD" Segmentation (2/2)

- For the case of an image consisting of several homogeneous color regions, the color classes are more separated from each other and the value of J is large.
- On the other hand, if all color classes are uniformly

distributed over the entire image, J tends to be small.

+	+	+	+	+	0	0	0	0
+	+	+	+	+	0	0	0	0
+	+	+	+	+	0	0	0	0
+	+	+	+	+	0	0	0	0
+	+	+	+	+	0	0	0	0
+	+	+	+	+	*	*	*	*
+	+	+	+	+	*	*	*	*
+	+	+	+	+	*	*	*	*
+	+	+	+	+	*	*	*	*

segmented class-map 1
 $J_+ = 0, J_* = 0, J_o = 0$
 $\bar{J} = 0$

+	+	+	+	+	*	0	*	0
+	+	+	+	+	0	*	0	*
+	+	+	+	+	*	0	*	0
+	+	+	+	+	0	*	0	*
+	+	+	+	+	*	0	*	0
+	+	+	+	+	*	0	*	0
+	+	+	+	+	0	*	0	*
+	+	+	+	+	*	0	*	0
+	+	+	+	+	0	*	0	*

segmented class-map 3
 $J_+ = 0, J_{\{*, o\}} = 0.011$
 $\bar{J} = 0.05$

Average $\bar{J}$ Notation

- Now, let's recalculate J over each segmented region instead of the entire class-map and define the average $\bar{J}$ by

$$\bar{J} = \frac{1}{N} \sum_k M_k J_k$$

- We propose $\bar{J}$ as the criterion to be minimized over all possible ways of segmenting the image given the number of regions.
- A "better" segmentation tends to have a lower value of $\bar{J}$. If the segmentation is good, each segmented region contains a few uniformly distributed color class labels and the resulting J value for that region is small. Therefore the overall $\bar{J}$ is small.

J-Image Introduced

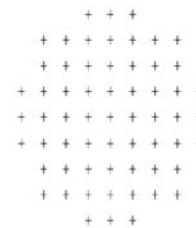
- Observe J , if applied to a local area of the class-map, is also a good indicator of whether that area is in the region interiors or near region boundaries. We can now think of constructing an image whose pixel values correspond to these J values calculated over small windows centered at the pixels
- The higher the local J value is, the more likely that the corresponding pixel is near a region boundary. The J -image is like a 3D terrain map containing valleys and hills that actually represent the region interiors and region boundaries, respectively.

Spatial Segmentation Algorithm

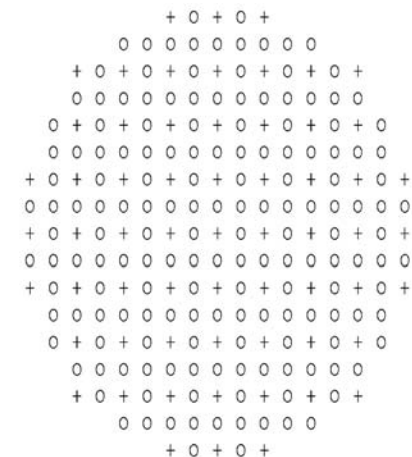
- The size of the local window determines the size of image regions that can be detected. Windows of small size are useful in localizing the intensity/color edges, while large windows are useful for detecting texture boundaries.
- The characteristics of the J-images allow us to use a region-growing method to

segment the image.

- From scale 1, the window size is doubled each time to obtain the next larger scale.

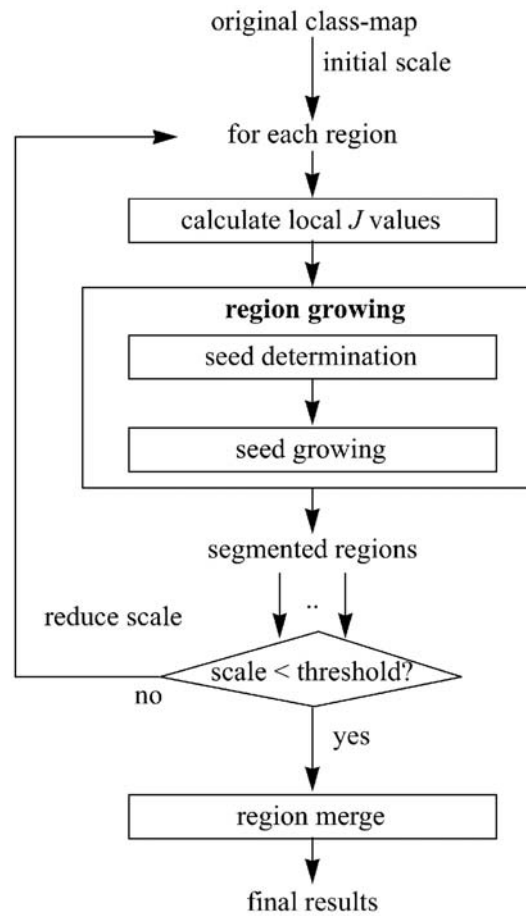


(a)



(b)

Flow-chart of Segmentation



- Consider the original image as one initial region. The algorithm starts the segmentation of the image at a coarse initial scale. Then, it repeats the same process on the newly segmented regions at the next finer scale.
- Region growing consists of determining the seed points and growing from those seed locations. Region growing is followed by a region merging operation to give the final segmented image.
- The user specifies the number of scales needed for the image, which affects how detail the segmentation will be.

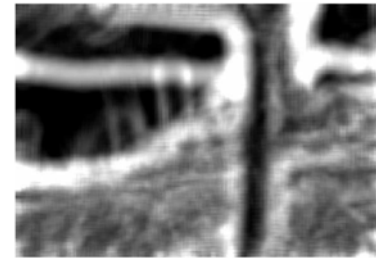
Segmentation Process



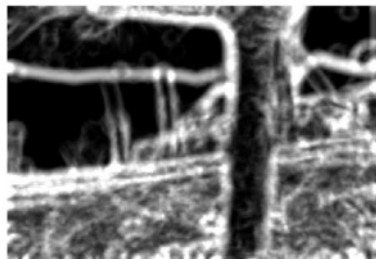
(a)



(b)



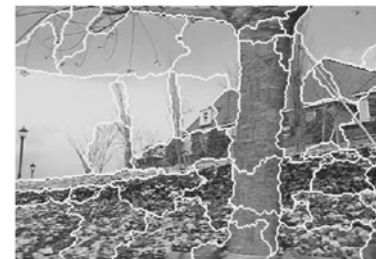
(c)



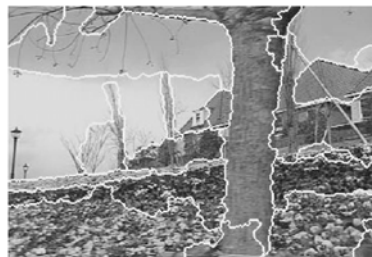
(d)



(e)



(f)



(g)

size: 352 x 240

original (a): $\bar{J} = 0.435$

scale 3 (e): $\bar{J} = 0.103$

scale 2 (f): $\bar{J} = 0.125$

final (g): $\bar{J} = 0.088$

Seed Determination

A set of initial seed areas are determined to be the basis for region growing. These areas correspond to minima of local J values.

scale	window (pixels)	sampling (1 / pixels)	region size (pixels)	min. seed (pixels)
1	9 x 9	1 / (1 x 1)	64 x 64	32
2	17 x 17	1 / (2 x 2)	128 x 128	128
3	33 x 33	1 / (4 x 4)	256 x 256	512
4	65 x 65	1 / (8 x 8)	512 x 512	2048

Heuristic Measure

- Calculate the average and the standard deviation of the local J values in the region, denoted by μ_J and T_J , respectively.
- Set a threshold T_J at

$$T_J = \mu_J + a\sigma_J$$

Pixels with local J values less than T_J are considered as candidate seed points. Connect the candidate seed points based on the 4-connectivity and obtain candidate seed areas.

- If a candidate seed area has a size larger than the minimum size listed in previous page, it is determined to be a seed.

Region Merge

- Oversegmented regions are merged based on their color similarity and the distance between two color histograms i and j is calculated by

$$D_{CH}(i, j) = \|P_i - P_j\|$$

Where P denotes the color histogram vector.

Spatiotemporal Segmentation Scheme

- The goal is to decompose the video into a set of objects in the spatiotemporal domain. Each object contains a homogenous color-texture pattern.
- It can be seen that the overall approach for each video frame is similar to the image segmentation work with the exception of seed tracking and post-processing procedures.
- The segmentation is performed on a group of consecutive frames. The number of frames, P , in each group can be set equal to the video shot length.

Seed Tracking

- The seeds in the first frame are assigned as initial objects.
- For each seed in the current frame, if it overlaps with an object from the previous frame, it is assigned to that object; Else a new object is created starting from this seed.
- If a seed overlaps with more than one object, these objects are merged.
- Repeat Steps 2 and 3 for all the frames in the group.
- A minimum time duration is set for the objects. Very short-length objects are discarded.

J_t Notation (1/3)

- Sometimes, false merges occur when two seeds of neighboring regions overlap with each other across the frames due to object motion.
- For two pixels at the same spatial location but in two consecutive frames, a modified measure J can be used to detect if their local neighbors have similar color distributions. This new measure is denoted as J_t

$$m_i = \frac{1}{N_i} \sum_{z \in Z_i} t_z$$

J_t Notation (2/3)

- Let

$$S_T = \sum_{z \in Z} \|t_z - m\|^2$$

and

$$S_W = \sum_{i=1}^C S_i = \sum_{i=1}^C \sum_{z \in Z_i} \|t_z - m_i\|^2$$

- The measure J_t is defined as

$$J_t = \frac{S_T - S_W}{S_W}$$

J_t Notation (3/3)

- When a region and its surroundings are static, the J_t values are small for all the points. When a region undergoes motion, the J_t value will become large when two points at the same spatial locations in the two frames do not belong to the same region.
- J_t is calculated for each point in the seeds. A threshold is set such that only points with small J_t values are used for tracking and the rest are discarded.
- After seed tracking, seed growing is performed on individual frames to obtain the final segmentation results.

Parameter Selection

- Threshold for the color quantization process. It determines the minimum distance between two quantized colors
- The number of scales desired for the image.
- The last parameter is the threshold for region merging.

Conclusion

- A new approach called JSEG is presented for the fully unsupervised segmentation of color-texture regions in images and video.
- A criterion for "good" segmentation is proposed. Applying the criterion to local image windows results in J-images, which can be segmented using a multiscale region growing method.
- For video, region tracking is embedded into segmentation. Results show the robustness of the algorithm on real images and video.