# Multi-label Image Segmentation via Point-wise Repetition

Gang Zeng
ETH Zürich
zengg@vision.ee.ethz.ch

Luc Van Gool
ETH Zürich / K.U. Leuven
vangool@vision.ee.ethz.ch

## Abstract

*Bottom-up segmentation tends to rely on local features. Yet, many natural and man-made objects contain repeating elements. Such structural and more spread-out features are important cues for segmentation but are more difficult to exploit. The difficulty also comes from the fact that repetition need not be perfect, and will actually rather be partial, approximate, or both in most cases. This paper presents a multi-label image segmentation algorithm that processes a single input image and efficiently discovers and exploits repeating elements without any prior knowledge about their shape, color or structure. The algorithm spells out the interplay between segmentation and repetition detection.*

*The key of our approach is a novel, point-wise concept of repetition. This is defined by point-wise mutual information and locally compares certain neighborhoods to accumulate evidence. This point-wise repetition measure naturally handles imperfect repetitions, and the parts with inconsistent appearances are recognized and assigned with low scores. An energy functional is proposed to include the point-wise repetition into the image segmentation process, which takes the form of a graph-cut minimization. Real scene images demonstrate the ability of our algorithm to handle partial and approximate repetition.*

## 1. Introduction

Multi-label image segmentation is an important task in computer vision, which provides crucial information for high level applications, such as object recognition, video tracking, scene reconstruction, image compression and so on. Two typical difficulties are: 1) the selection of image features and 2) feature clustering for label assignment.

Color and texture are popular features, with several outstanding examples in the literature [28, 8]. Humans definitely take a more holistic approach to segmentation though, as is already known from the Gestalt school. High level regularities in the pattern seem to jump into the eye. If we look around, we cannot help but to notice repetitive patterns. Many natural and man-made objects exhibit repetition of some kind. Fruits and animals have regularities in their skins, buildings have repeating windows, clothes often have periodic patterns, etc. Such repetitions bind together groups of pixels that might be widely spread out over the image and may even be disjoint. Therefore, observed repetitions are a crucial asset for segmentation, especially at the stage of the feature clustering.

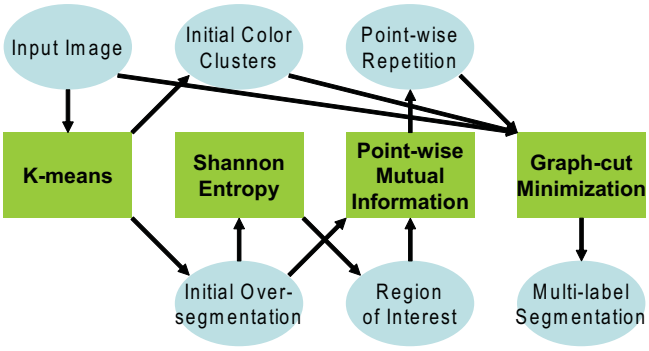



Figure 1. Overview of the four stages of the proposed algorithm.

However, from a computational point of view, repetition detection is not an easy task. This is all the more the case since repetition often is partial, approximate or both. Without a clear premonition about the repeating element's shape and size, there are many degrees of freedom. Moreover, element appearance may differ somewhat from element to element, e.g. because of different illumination, viewpoint, natural variation, etc. These difficulties have hampered the widespread use of repetition detection in segmentation.

### 1.1. Overview

We propose a multi-label image segmentation algorithm that processes a single input image and efficiently discovers and exploits repeating elements without any prior knowledge about their shape, intensity or structure. A new formulation is proposed to describe the interplay between segmentation and repetition detection. The former produces labels for the latter, while the latter in turn provides corresponding points for the former. An overview of the four stages of the proposed algorithm is shown in Fig. 1.

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

At the first stage, we use the k-means method to produce initial color clusters and the corresponding initial over-segmentation of the image. Through their colors, pixels are regarded as 3D points in the CIE $L^*u^*v^*$ color space. Compared with the RGB, CMYK and HSV color spaces, CIE $L^*u^*v^*$ is designed to approximate human vision, where a good metric for assessing perceptual differences among colors is given by the simple Euclidean distances [13]. Voronoi cells in this color space yield the labels of this segmentation, i.e. the clusters of object colors. Please note that the initial over-segmentation does not have to be accurate, and the labels can later be split and merged by optimizing our energy functional. The k-means method is initialized to have 64 clusters in our experiments. The implementation details of this first stage can be found in Sec. 4.1.

At the second stage, with the initial over-segmentation at hand, we compute the *Shannon entropy* for the segmentation labels in the local neighborhood of each pixel. Low values indicate pixels in color-homogeneous areas. Their detection will bring appreciable time savings in the sequel of our pipeline. This second stage is described in Sec. 3.1.

The third stage is the key step of our algorithm. Unlike the previous approaches where repetition detection is based on the comparison among entire regions in at least part of the pipeline, we propose a *point-wise* repetition detection scheme. The repeating regions are then collected as the maximal areas inside which point-wise repetition is prominent. The scheme can cope with imperfect repetitions, where deviations are assigned low scores. The formulation and discussion of this third stage are found in Sec. 3.2.

At the last stage, based on the above formulation of the interplay between segmentation and repetition detection, an energy functional is designed to describe the multi-label image segmentation problem. It encodes not only the point-wise repetition scores but also the color homogeneity and the expectations embodied by smoothness priors. A global solution is found through a graph-cut minimization process. The labels of the initial over-segmentation are split and merged, and their boundaries are refined. The functional and implementation of the last stage are the subjects of Sec. 3.3.

### 1.2. Related Work

**Image Segmentation** In the context of multi-label image segmentation from a single image, there are two main streams of research in the literature: feature-space based and image-domain based methods. The idea behind feature space based methods is that color (or possibly texture) is a constant feature of objects' surfaces so that in a certain feature space it forms a distinguishable cluster. Among the most common algorithms of cluster analysis is the k-means algorithm and its variations [21, 17]. Histogram thresholding methods [9, 20] have mainly been designed for gray-level images.

Image-domain based methods take the spatial compactness into consideration. The class of regions these algorithms return is expected to be homogeneous with respect to the feature space and compact in the image domain. Strategies include split&merge [26], region-growing [25] and graph-theoretical methods [27, 24]. The proposed algorithm belongs to the image-domain methods, where we employ k-means for initialization but a graph-cut process for optimization.

**Texture Regularity Detection** Our work is inspired by previous research on texture regularity, which has applications such as shape-from-texture [10, 6] and texture segmentation [22, 19]. Finding the basic texture units (texels) in real images has proven a difficult task, sometimes circumvented through user interaction [16]. An automatic texel detection method has been proposed by Hays *et al*. [11]. It is based on the discovery of the lattices of near-regular textures. The lattice identifies the texels as well as their spatial organisation. Compared with this work, our repetition detection does not assume a lattice layout, and handles repeated elements even if they are disjoint in the image domain.

**Energy Minimization** Another important class of related works is based on energy formulations which are minimized via discrete optimization techniques. The pioneering graph cuts technique [2] addresses the foreground/background interactive segmentation in still images via max-flow/min-cut energy minimization. These graph based methods were made popular first through the Normalized Cut formulation of [24] and more recently by the energy minimization method of [3].

There has also been some previous work which adds high-level information to the benefit of more general low-level segmentation methods [15, 5, 7, 23]. In contrast, the algorithm proposed here employs more generic repetition (regularity) criteria instead of specific shape priors.

## 2. Segmentation-repetition Formulations

Given an input color image $I(\mathbf{x})$, the goal is to segment it into different labels $L(\mathbf{x})$, which represent different types of elements. The value of $L(\mathbf{x})$ is an integer, representing the label of the current pixel, and thus all corresponding pixels of the same type of elements can ideally be detected as $Label_i = \{\mathbf{x}|L(\mathbf{x}) = i\}$.

This is improbable to happen automatically when segmenting a scene purely on the basis of color or texture. The core of our approach to multi-label segmentation lies in adding a role for repetition. Suppose we have detected a pair of repeating regions $(R_l, R_r)$ (found starting at the pixel level though, as will be explained). The system then

assumes that pairs of corresponding points $(\mathbf{x}_l, \mathbf{x}_r)$ between $R_l$ and $R_r$ should get the same label, *i.e.* $L(\mathbf{x}_l) \equiv L(\mathbf{x}_r)$. This is the basic constraint provided by repetition.

In the following sections we first assume the prior availability of either repetition or segmentation, to investigate how that would help the other component. In the remainder of the paper, the two processes will then be made integral parts of one and the same scheme though.

## 2.1. Known Repetition

Suppose repeating regions $\{(R_l, R_r)\}$ have been provided. Based on the basic constraint proposed in the previous paragraphs, an energy functional for segmentation can be defined as follows:

$$\begin{aligned} E_R &= E_{data} + \alpha_E E_{repR} + \beta_E E_{prior} \\ E_{data} &= \int_{\mathbf{x}} |I(\mathbf{x}) - I_{L(\mathbf{x})}|^2 d\mathbf{x} \\ E_{repR} &= \sum_{(R_l,R_r)} \int_{\mathbf{x}_l \in R_l} \bar{\delta}(L(\mathbf{x}_l) - L(\mathbf{x}_r)) d\mathbf{x}_l \\ E_{prior} &= \int_{\mathbf{x}} \sum_{\mathbf{x}_N \in N_{\mathbf{x}}} \bar{\delta}(L(\mathbf{x}) - L(\mathbf{x}_N)) g(|I(\mathbf{x}) - I(\mathbf{x}_N)|) d\mathbf{x} \\ \text{with} \quad \bar{\delta}(x) &= \begin{cases} 0 & \text{if } x = 0; \\ 1 & \text{otherwise.} \end{cases} \end{aligned} \tag{1}$$

Here $I_{L(\mathbf{x})}$ is chosen from the set of initial colors $\{I_i\}$ selected by the color-based over-segmentation. As with median filtering, we will stick to this initial set. Some colors may disappear through further optimizations, but no new colors are added to this initial color set. $N_{\mathbf{x}}$ denotes the 4-neighborhood for $\mathbf{x}$. $g(|I(\mathbf{x}) - I(\mathbf{x}_N)|)$ is a decreasing function with the local color variation. Its role is to accept label changes, but only near the boundaries of color discontinuities.

This cost functional penalises the variations in color for each label through the first term $E_{data}$, and differences in labels for any pair of corresponding pixels between repeating regions through the second term $E_{repR}$, and also changes in labels throughout the image except at color boundaries through the third term $E_{prior}$.

The problem with the above energy functional is that it takes an all-or-nothing approach to repetition. Either pixels are in "repeating regions" or not. In reality, repetition is more gradual in nature, with some parts being repeated more faithfully than others. Pixels where the exactitude of repetition is less perfect, should be made to weigh less in terms of cost, although the entire regions they're in may still look similar. This served as a motivation to propose *point-wise repetition detection*, and together with it a repetition measure $s(\mathbf{x}_l, \mathbf{x}_r)$ (see Sec. 3.2).

Once we have this point-wise repetition measure $s(\mathbf{x}_l, \mathbf{x}_r)$, the above energy functional can be improved as follows:

$$\begin{aligned} E &= E_{data} + \alpha_E E_{rep} + \beta_E E_{prior} \\ E_{rep} &= \sum_{(\mathbf{x}_l,\mathbf{x}_r)} \bar{\delta}(L(\mathbf{x}_l) - L(\mathbf{x}_r)) s(\mathbf{x}_l, \mathbf{x}_r) d\mathbf{x}_l. \end{aligned} \tag{2}$$

## 2.2. Known Segmentation Label

Before dealing with the integrated handling of segmentation and repetition detection, we first also look at the opposite case: repetition detection with a label function $L(\mathbf{x})$ already available from segmentation.

Given $L(\mathbf{x})$, let $p(X)$ and $p(Y)$ denote the probability distributions of the labels $X$ and $Y$ within regions $R_l$ and $R_r$ respectively. $p(X = i)$ and $p(Y = j)$ and their joint probability distribution function $p(X = i, Y = j)$ are thus defined as follows:

$$\begin{aligned} p(X = i) &= \frac{\int_{\mathbf{x}_l \in R_l, L(\mathbf{x}_l) = i} d\mathbf{x}_l}{\int_{\mathbf{x}_l \in R_l} d\mathbf{x}_l} \\ p(Y = j) &= \frac{\int_{\mathbf{x}_r \in R_r, L(\mathbf{x}_r) = j} d\mathbf{x}_r}{\int_{\mathbf{x}_r \in R_r} d\mathbf{x}_r} \\ p(X = i, Y = j) &= \frac{\int_{\mathbf{x}_l \in R_l, L(\mathbf{x}_l) = i, L(\mathbf{x}_r) = j} d\mathbf{x}_l}{\int_{\mathbf{x}_l \in R_l} d\mathbf{x}_l}, \end{aligned} \tag{3}$$

where $(\mathbf{x}_l, \mathbf{x}_r)$ denote corresponding positions.

With the assumption that $L(\mathbf{x})$ has been properly determined in the image, if $R_l$ and $R_r$ are repeating regions, we should have $L(\mathbf{x}_l) \equiv L(\mathbf{x}_r)$. The repetition measure $S_R(R_l, R_r)$ can be defined as

$$S_R(R_l, R_r) = \sum_i p(X = i, Y = i) \in [0, 1]. \tag{4}$$

When repetition is partial or approximate, some parts in two repeating regions may have inconsistent appearances *i.e.* $L(\mathbf{x}_l) \neq L(\mathbf{x}_r)$. Thus, repetition detection can rather be based on searching all pairs of regions $(R_l, R_r)$ with $S_R(R_l, R_r) > \tau \approx 1$, where $\tau$ is a threshold to control the tolerance regarding inconsistent appearances. As a further remark on this issue, once we have the point-wise repetition measure $s(\mathbf{x}_l, \mathbf{x}_r)$ (see Sec. 3.2), point-wise repetition detection can be performed in a similar manner.

# 3. Segmentation via Point-wise Repetition

In general, when there is no prior information about repetition or segmentation labels, it is less obvious how the system may bootstrap itself out of that situation. Typical solutions will start from a reasonable, initial guess and then try to iteratively converge to a solution. Here we propose a single-step procedure. Our initial guess for the labels still corresponds to the initial color-based over-segmentation, as introduced in Sec. 1.1.

As we can no longer assume that corresponding pixels will have the same label, a more robust repetition measure is due, that can still drive the process ahead under such conditions. *Mutual Information* (MI) [4] is such a measure, as it is very tolerant to the kind of mapping that exists between random variables, as long as such mapping is relatively systematic [18]. MI of two random variables $X$ and $Y$ is extracted from their *Shannon entropies* $H(X)$ and $H(Y)$, and their *joint entropy* $H(X,Y)$.

$$\begin{aligned} H(X) &= -\sum_i p(X=i)\log p(X=i) \\ H(Y) &= -\sum_j p(Y=j)\log p(Y=j) \\ H(X,Y) &= -\sum_{i,j} p(X=i,Y=j)\log p(X=i,Y=j) \\ MI(X,Y) &= \sum_{i,j} p(X=i,Y=j)\log\frac{p(X=i,Y=j)}{p(X=i)p(Y=j)} \\ &= H(X)+H(Y)-H(X,Y). \end{aligned} \tag{5}$$

While the Shannon entropy measures the uncertainty associated with a random variable, MI measures how much information one variable contains about the other, i.e. the mutual dependence between the two variables.

With two repeating regions, labels in either region almost are a deterministic function of those in the other region, and mutual information has a high score. We thus can define the repetition measure $S(R_l, R_r)$ between two repeating regions based on the mutual information as follows:

$$S(R_l, R_r) = MI(X,Y). \tag{6}$$

### 3.1. Region of Interest

However one looks at segmentation, i.e. either as finding region boundaries or as delineating homogeneous areas, regions with and without variation in the basic underlying feature(s) need to be treated differently. We bring the entropy of the initial label image to bear (i.e. for the over-segmented initialization) and only calculate point-wise MI (and include graph cut links) for pairs of pixels that both lie in high entropy regions. The reason is that MI is bounded from above by the lowest of the two entropies. Large entropies are found around the boundaries of color discontinuities. Fig. 4(b) shows two examples of the difference in entropy between homogeneous and boundary areas. Thus, edge regions yield the strongest indications about repetition. Restricting MI scores for such pixel pairs also reduces computation times.

### 3.2. Point-wise Repetition

The previously defined repetition measures $S_R(R_l, R_r)$ and $S(R_l, R_r)$ in Eqns. (4)&(6) are both region based.

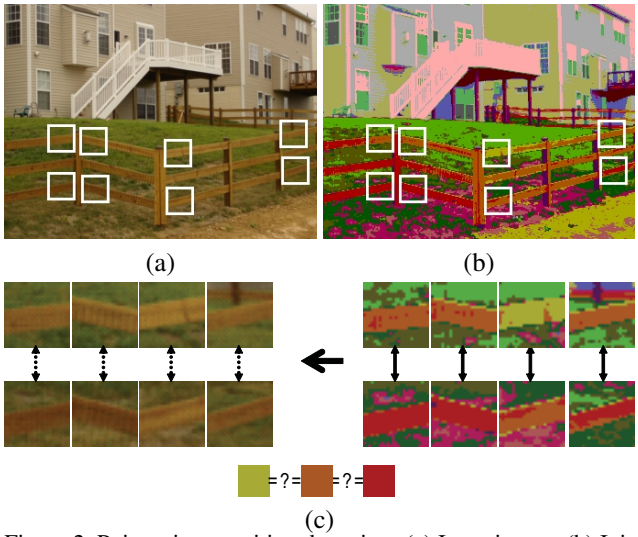



Figure 2. Point-wise repetition detection: (a) Input image; (b) Initial over-segmentation by k-means; (c) Different labels are connected through several pairs of point-wise repetition.

However, in practice real objects often exhibit partial or approximate repetition (see Fig. 2(a)). The region-based measures would either fail to recognize repetition or gloss over the subregions where such deviations occur. This motivates us to go after *point-wise repetition detection* and define a related repetition measure $s(x_l, x_r)$.

What we ideally want is a more local version of MI. This is delivered by the so-called *point-wise MI*:

$$s(\mathbf{x}_l, \mathbf{x}_r) = \log\frac{p(X=L(\mathbf{x}_l), Y=L(\mathbf{x}_r))}{p(X=L(\mathbf{x}_l))p(Y=L(\mathbf{x}_r))}. \tag{7}$$

It is based on the probabilities $p(X=L(\mathbf{x}_l), Y=L(\mathbf{x}_r))$, $p(X=L(\mathbf{x}_l))$ and $p(Y=L(\mathbf{x}_r))$ instead of the whole distribution functions $p(X,Y)$, $p(X)$ and $p(Y)$. The fractions of (foreground) labels that contain the current pair of pixels are compared, while the other (background) labels are not considered. In order to estimate these probabilities, we sample from a square neighborhood around $\mathbf{x}_l$ and $\mathbf{x}_r$. The diameter of the squares is given by a pre-defined parameter.

This point-wise repetition measure is consistent with the previously defined $S(R_l, R_r)$ (see Eqn. (6)): when summing all point-wise MI values of pixels in a region, one arrives back at this original value $S(R_l, R_r)$ (after division by the region's area), *i.e.*

$$\begin{aligned} & S(R_l, R_r) \\ &= \sum_{i,j} p(X=i,Y=j)\log\frac{p(X=i,Y=j)}{p(X=i)p(Y=j)} \\ &= \sum_{i,j} \frac{\int_{\mathbf{x}_l\in R_l, L(\mathbf{x}_l)=i, L(\mathbf{x}_r)=j} d\mathbf{x}_l}{\int_{\mathbf{x}_l\in R_l} d\mathbf{x}_l} \log\frac{p(X=i,Y=j)}{p(X=i)p(Y=j)} \\ &= \frac{\int_{\mathbf{x}_l\in R_l} s(\mathbf{x}_l,\mathbf{x}_r)d\mathbf{x}_l}{\int_{\mathbf{x}_l\in R_l} d\mathbf{x}_l}. \end{aligned} \tag{8}$$

This allows a region-growing process, which starts from a pair of corresponding points with a high point-wise repetition score and propagates in their neighborhood to form a pair of maximal repeating regions.

Especially when given an over-segmentation, the same type of objects may belong to different labels (see Fig. 2(b)), due to different illumination, viewpoint, natural variation, etc. On the other hand, these differences in turn yield high penalty scores of $E_{rep}$ in Eqn. (2). Minimizing the total energy $E$ trades off initial over-segmentation (through $E_{data}$) vs detected repetition (through $E_{rep}$) and smoothness. The changes between the initial and final labels indicates the label split and merge operations. The final labels without support (no longer used for any pixel) are detected and removed, and thus the number of clusters provided by the initial over-segmentation is reduced.

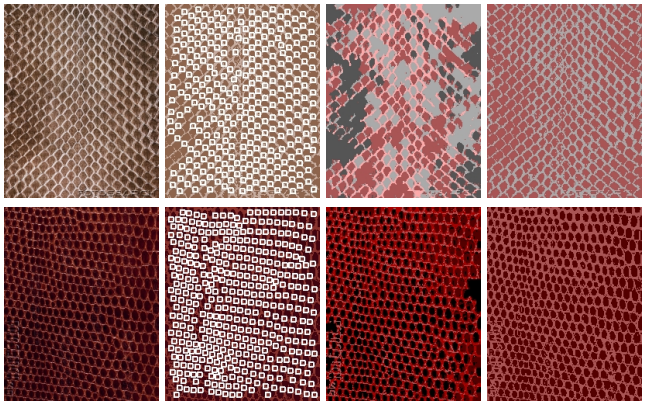
### 3.3. Energy Minimization

With the point-wise repetition measure $s(\mathbf{x}_l, \mathbf{x}_r)$ defined in Eqn. (7) plugged into the energy functional $E$ in Eqn. (2), we have arrived at the functional that we try to minimize. For this minimization, we embed our problem into a graph, and use the classic max-flow/min-cut algorithm. Kolmogorov and Zabih [14] give a characterization of which energy functions can be minimized using graph-cuts, and they also provide a graph-construction method. Readers are referred to their paper for more detailed information.

In the following paragraphs, we follow their approach and focus on the proof that validates our energy minimization problem, *i.e.* we convert the energy functional in Eqn. (2) into a binary form which is graph-representable. The conversion is done by the $\alpha-$expansion operation: Any configuration $L_\alpha(\mathbf{x})$ within a single $\alpha-$expansion of the initial configuration $L(\mathbf{x})$ can be encoded by a binary function

$$\triangle L(\mathbf{x}) = \begin{cases} 0, & \text{if } L_\alpha(\mathbf{x}) = L(\mathbf{x}); \\ 1, & \text{if } L_\alpha(\mathbf{x}) = \alpha. \end{cases} \tag{9}$$

The algorithm runs through all values for $\alpha$ in turn, with the graph-cut only being allowed to change labels to that $\alpha$ value at the time. Let $L_\triangle(\mathbf{x})$ denote a configuration defined by $\triangle L(\mathbf{x})$. Then, we have the energy of binary variables, $\triangle E(\triangle L) = E(L_\triangle) - E(L)$. It can be proven that this binary form is graph-representable. Intuitive proofs are that $\bar{\delta}(L(\mathbf{x}_l) - L(\mathbf{x}_r)) = 0$ when $\triangle L(\mathbf{x}_l) = \triangle L(\mathbf{x}_r) = 1$ (*i.e.* $L(\mathbf{x}_l) = L(\mathbf{x}_r) = \alpha$), and that $\bar{\delta}(L(\mathbf{x}) - L(\mathbf{x}_N)) = 0$ when $\triangle L(\mathbf{x}) = \triangle L(\mathbf{x}_N) = 1$ (*i.e.* $L(\mathbf{x}) = L(\mathbf{x}_N) = \alpha$). Due to space limitation, the proof and implementation details are omitted, but they will be provided in a technical report.



(a) (b) (c) (d)

Figure 3. Segmentation results: (a) Input image; (b) A group of corresponding points by point-wise repetition detection (shown in white squares); (c) Segmentation without repetition detection; (d) The proposed segmentation via point-wise repetition.

## 4. Implementation and Experimental Results

### 4.1. Implementation Details

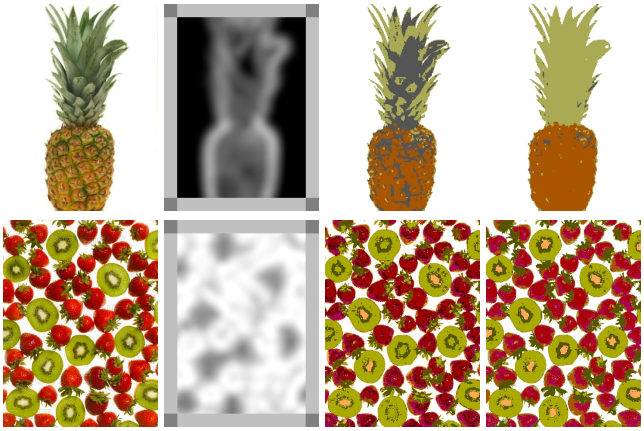
In all our experiments, the k-means method is initialized with 64 clusters. For computing $s(\mathbf{x}_l, \mathbf{x}_r)$ we use $64 \times 64$ windows to estimate the probabilities $p(X = L(\mathbf{x}_l))$, $p(Y = L(\mathbf{x}_r))$ and $p(X = L(\mathbf{x}_l), Y = L(\mathbf{x}_r))$. A larger size introduced a higher computational cost, while a smaller size provided insufficient samples. The weights of the energy functional in Eqn. (2) are fixed to $\alpha_E = 1$ and $\beta_E = 0.25$.

Finally, in order to accelerate the whole process, we limit the point-wise repetition detection within the region of interest, where the entropy score is high, *i.e.* $H(X) \geq 0.5$. We also rescale all input images such that the number of pixels equals that of a video frame, *i.e.* $640 \times 480$. $O(n^2)$ point-wise matching and repetition detection in the third stage would yield low efficiencies for larger images. A multi-resolution approach is envisaged, but has not been implemented yet.

### 4.2. Results and Discussion

We show experimental results for several challenging, real-world images. We compare segmentations without and with the use of repetition (i.e. without and with the $E_{rep}$ term in Eqn. (2)).

Fig. 3(a) shows repetitive animal skins. As can be seen, the repetitions are far from perfect. Without regularity information, the graph-cut minimization [3] either ignores some thin boundaries or assigns multiple labels to the same patterns in Fig. 3(c). With point-wise repetition detection integrated (a group of corresponding points is shown as



(a) (b) (c) (d)

Figure 4. Segmentation results: (a) Input image; (b) Shannon entropy with high values near edges; (c) Segmentation without repetition detection; (d) The proposed segmentation via point-wise repetition.

white squares in Fig. 3(b)), the segmentations are far cleaner in Fig. 3(d). The running time without and with repetition detection for these two examples are about 150 and 300 seconds respectively on a Pentium4 3.2GHz machine.

Fig. 4(a) shows a second experiment. Local Shannon entropies are displayed in gray color in Fig. 4(b), with high values near edges and low values in color homogeneous areas. As mentioned, in order to gain efficiency we limit the point-wise repetition detection to regions of interest, where $H(X) \geq 0.5$. The pineapple exhibits regularities both in the arrangement of its leaves and on its shell. Compared to the color-only segmentation, the repetition-enabled segmentation manages to detect these two parts (see Figs. 4(b)&(c)). Similarly, in the strawberry + kiwis case, the segmentation based on repetition is far more consistent throughout the different instances (e.g. kiwi hearts are systematically placed into the same segment class). Though the granularity of the segmentations can be influenced through the algorithm's parameters, it is crucial that the repetition-enabled segmentation yields much more consistent results. The running times are about 100 and 250 seconds for these two examples.

Fig. 5(a) shows some interesting outdoor scenes. Fences are an interesting case, as the repeated parts are interleaved with unrelated portions of the background. Yet, one would like to see the fence come out as one segment. The need for repetition detection is illustrated from the results in Fig. 5(b)&(c), where the fence is split up into different segments, as the color-based segmentation gets confused by similar ground colors. Fig. 5(d) shows the repetition-based segmentation. As this and the other examples in the figure show, the algorithm constantly segments the fences out, with clear boundaries even if the gaps are rather thin and backgrounds show quite some variation. The running time for these examples are about $250 - 300$ seconds.

As a last example, also most buildings contain repeating substructures. This said, the repetition can be partial (not the entire pattern is repeated, or copies are not entirely visible), approximate (the copies deviate in their appearance), or both. As a common example, some windows are closed, while others are open or partially open. Window blinds, shadows, reflections, and curtains all result in different appearances. Typical examples are shown in Fig. 6(a). Such variations easily lead color-based segmentation astray (see Fig.6(b)). Via point-wise repetition, the walls, roofs and roads are all segmented much more systematically into single clusters (see Fig.6(c)). Even some detailed window structures come out of our algorithm. The running time of the whole process for these facades is about $300 - 400$ seconds.

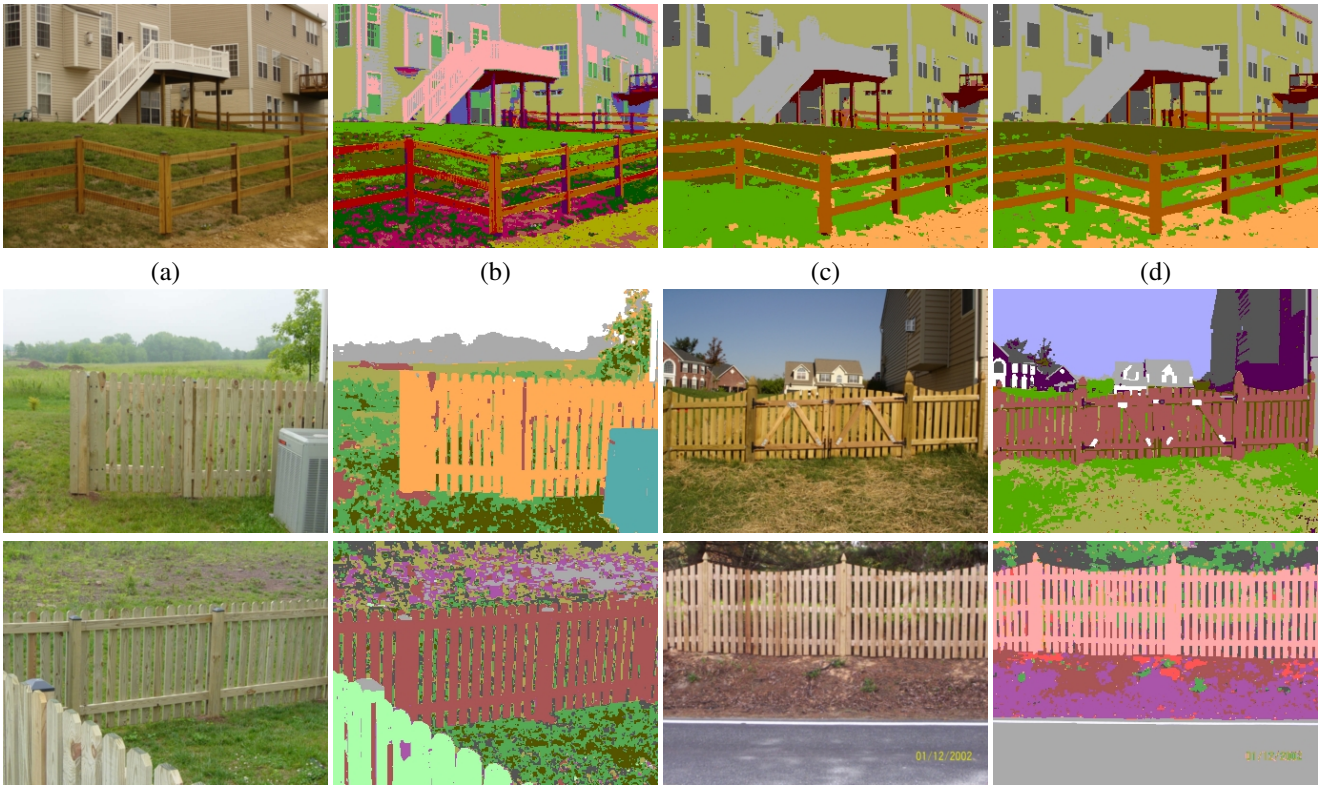
The main computation time goes to point-wise repetition detection, for which times go up with the square of the number of pixels. We have limited the search to regions with high entropies, thereby containing the problem. In future work, we plan to exploit the regular distribution of repetition to further accelerate the algorithm. A top-down hierarchical data structure will be employed to first search at a global level and to then add details at each lower level.

## 5. Conclusion

We have proposed an automatic multi-label segmentation algorithm, which uses a single input image and exploits the interplay between segmentation and repetition detection to solve the ambiguities. It is based on a point-wise repetition detection to split and merge an initial over-segmentation and an energy minimization formulation that combines this repetition detection with a preference for color homogeneity and the expectations embodied by smoothness priors. The corresponding problem is optimized via a graph-cut minimization process. Real scene images demonstrate the ability of our algorithm to handle partial and approximate repetition. Compared with previous methods, ours does not rely on any user interaction and yet can produce accurate segmentations at pixel level.

## Acknowledgments

The authors gratefully acknowledge support by the European Commission, through IST project Cyberwalk, and the Flemish Fund for Scientific Research, FWO. The authors also thank Pascal Müller and Filip Van den Borre for their constructive comments on improving this paper.



(a) (b) (c) (d)

(a) (d) (a) (d)

Figure 5. Segmentation results: (a) Input image; (b) Initial over-segmentation by k-means; (c) Segmentation without repetition detection; (d) The proposed segmentation with point-wise repetition.

## References


[1] *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), 20-26 June 2005, San Diego, CA, USA*. IEEE Computer Society, 2005.

[2] Y. Boykov and M.-P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. In *ICCV*, pages 105–112, 2001.

[3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001.

[4] T. M. Cover and J. A. Thomas. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, 2006.

[5] D. Cremers. Dynamical statistical shape priors for level set-based tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(8):1262–1273, 2006.

[6] D. A. Forsyth. Shape from texture without boundaries. In Heyden et al. [12], pages 225–239.

[7] D. Freedman and T. Zhang. Interactive graph cut based segmentation with shape priors. In *CVPR (1)* [1], pages 755–762.

[8] J. Freixenet, X. Muñoz, D. Raba, J. Martí, and X. Cufí. Yet another survey on image segmentation: Region and boundary information integration. In Heyden et al. [12], pages 408–422.

[9] K. S. Fu and J. K. Mui. A survey on image segmentation. *Pattern Recognition*, 13(1):3–16, 1981.

[10] J. Gårding. Surface orientation and curvature from differential texture distortion. In *ICCV*, pages 733–739, 1995.

[11] J. Hays, M. Leordeanu, A. A. Efros, and Y. Liu. Discovering texture regularity as a higher-order correspondence problem. In A. Leonardis, H. Bischof, and A. Pinz, editors, *ECCV (2)*, volume 3952 of *Lecture Notes in Computer Science*, pages 522–535. Springer, 2006.

[12] A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors. *Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Copenhagen, Denmark, May 28-31, 2002, Proceedings, Part III*, volume 2352 of *Lecture Notes in Computer Science*. Springer, 2002.

[13] R. Hunt. *The Reproduction of Colour*. Wiley, 1967.

[14] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(2):147–159, 2004.

[15] M. P. Kumar, P. H. S. Torr, and A. Zisserman. Obj cut. In *CVPR (1)* [1], pages 18–25.

[16] Y. Liu, W.-C. Lin, and J. Hays. Near-regular texture analysis and manipulation. *ACM Trans. Graph.*, 23(3):368–376, 2004.

[17] J. Luo, R. T. Gray, and H.-C. Lee. Incorporation of derivative priors in adaptive bayesian color image segmentation. In *ICIP (3)*, pages 780–784, 1998.

(a) (b) (c)

Figure 6. Segmentation results: (a) Input image; (b) Segmentation without repetition detection; (c) The proposed segmentation with pointwise repetition.


[18] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging*, 16(2):187–198, 1997.

[19] J. Malik, S. Belongie, J. Shi, and T. K. Leung. Textons, contours and regions: Cue integration in image segmentation. In *ICCV*, pages 918–925, 1999.

[20] N. R. Pal and S. K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, 1993.

[21] S. H. Park, I. D. Yun, and S. U. Lee. Color image segmentation based on 3-d clustering: morphological approach. *Pattern Recognition*, 31(8):1061–1076, 1998.

[22] J. Puzicha, T. Hofmann, and J. M. Buhmann. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In *CVPR*, pages 267–272. IEEE Computer Society, 1997.

[23] M. Rochery, I. H. Jermyn, and J. Zerubia. Higher order active contours. *International Journal of Computer Vision*, 69(1):27–42, 2006.

[24] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8):888–905, 2000.

[25] A. Trémeau and N. Borel. A region growing and merging algorithm to color segmentation. *Pattern Recognition*, 30(7):1191–1203, 1997.

[26] J.-P. Wang. Stochastic relaxation on partitions with connected components and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(6):619–636, 1998.

[27] Z. Wu and R. M. Leahy. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 15(11):1101–1113, 1993.

[28] Y. J. Zhang. Evaluation and comparison of different segmentation algorithms. *Pattern Recognition Letters*, 18(10):963–974, 1997.