SWAT & Normalized Cuts.

Segmentation by Weighted Aggregation (SWA) is a hierarchical algorithm for segmenting images.

But it does not yield a complete segmentation into disjoint regions

\[ L = \bigcup_{i \neq j} L_i \quad \text{and} \quad L_i \cap L_j = \emptyset, \quad \text{if } i \neq j. \]

Instead it seeks to find segments in the image which are homogeneous (in terms of a criterion we will define later). These segments can overlap and will not necessarily cover the image.

SWA has several advantages.

1. Very efficient implementation - algorithm runs in seconds on a 200 x 300 image.

2. It processes the image at multiple scales and outputs a set of salient segments at all scales.

But determining which segments are salient and should be kept requires a second stage.
First we describe the basic algorithm.

Start with a grid $\Lambda^0$ -- the image lattice.

Define affinities $w_{ij}^0$ between pixels in $\Lambda^0$:

$$w_{ij}^0 = e^{-\gamma|I_i - I_j|}$$

where $I_i, I_j$ are the intensities at nodes $i, j \in \Lambda^0$.

(Do this for nearest neighbors in the lattice only)

Now, select a subset $\Lambda^1$ of nodes ($\Lambda^1 \subset \Lambda^0$) obeying the properties:

$$A \subseteq \Lambda^0 \setminus \Lambda^1 \quad (i.e. \text{ } i \in \Lambda^0 \text{ but } i \notin \Lambda^1)$$

we have

$$\sum_{k \in \Lambda^1} w_{ik}^0 \geq \beta \sum_{j \in \Lambda^0} w_{ij}^0$$

Intuitively, a pixel $i$ is left behind ($i \in \Lambda^0 \setminus \Lambda^1$) if it is represented by nodes in $\Lambda^1$, if it has sufficiently high affinity to nodes in $\Lambda^1$.

The nodes $\Lambda^1$ are representatives of $\Lambda^0$. 

The algorithm to select \( V^1 \) proceeds as follows (alternative algorithms are possible).

First, order the nodes in \( V^0 \).

Proceed sequentially.

If a node is weakly attached to current blocks of nodes, then create a new block containing this node.

Let \( C^{i-1} \) denote the set of blocks before testing node \( i \).

Check inequality:

\[
\max_{j \in C^{i-1}} w_{i,j} \geq \bar{z} \sum_{l \in C^{i-1}} w_{i,l} \quad (z \geq 0.1)
\]

If satisfied, set \( C^i = C^{i-1} \)

If not satisfied, set \( C^i = C^{i-1} \cup \{i\} \)

Note: this selection of \( V^1 \) depends on the initial ordering of the nodes. This does not seem to affect the performance.

Typically choose \( \bar{z} \in [0.1, 1.0] \).
Define an interpolation matrix \( p^o_{ik} \) for nodes left to hide.

\[
p^o_{ik} = \frac{\omega^o_{ik}}{\sum_{j \in \Lambda^1} \omega^o_{ij}} \quad k \in \Lambda^1, \quad i \in \Lambda^0 \cap \Lambda^1
\]

\[
p^o_{ii} = \begin{cases} 1 & i \in \Lambda^0 \cap \Lambda^1 \\ 0 & \text{otherwise} \end{cases}
\]

Intuitively, \( p^o_{ik} \) is the probability that node \( i \in \Lambda^0 \) is represented by node \( k \in \Lambda^1 \).

\[
\sum_{k \in \Lambda^1} p^o_{ik} = 1, \quad \forall i \in \Lambda^0
\]

Then define new affinities of \( \Lambda^1 \) by

\[
\omega_{ij} = \sum_{k \in \Lambda^0} \omega^o_{ik} \omega^o_{jk} p^o_{il}
\]

Then proceed as before using \( \Lambda^1 \) as the new grid and the \( \omega^o \) as the affinities.

\( \Lambda^0 > \Lambda^1 > \Lambda^2 > \Lambda^3 \)
\[ \omega_{nk}^\mu = \sum_{i,j \in \Lambda_{k-1}} p_{i,k}^{\mu-1} \omega_{ij}^{\mu-1} p_{j,k}^{\mu-1} \]

\[ p_{ik}^{\mu-1} = \frac{\omega_{ik}^{\mu-1}}{\sum_{j \in \Lambda_k} \omega_{jk}^{\mu-1}}, \quad k \in \Lambda_k. \]

How to justify this coarsening?

SWA was adapted from a method for implementing differential equations known as Algebraic Multi-Grid (AMG). AMG proposes a way to solve the differential equation on a hierarchy. The hierarchy is chosen by AMG and depends on the data in a specified manner. But the criteria for solving differential equations efficiently are different from those used to segment images—so much of the theory of AMG may be irrelevant for segmentation.

Note: the alternative to AMG, is Geometric Multigrid. What you use geometric rules (indeed of data) to generate the hierarchy (e.g., leave out every second pixel).
The concept of salience

Define a measure of salience for a segment $S$ (set of nodes)

\[ u_i^0 = \begin{cases} 1, & \text{if } i \text{ is in the segment} \\ 0, & \text{otherwise} \end{cases} \]

\[ \Gamma \{ u^0 \} = \sum_{ij \in \mathcal{E}} w_{ij} (u_i^0 - u_j^0)^2 \overline{w_{ij} u_i^0 u_j^0} \]

Here, \( \sum_{ij \in \mathcal{E}} w_{ij} (u_i^0 - u_j^0)^2 \) gives the cost of the affinity across the boundaries of the salient. \( \sum_{ij \in \mathcal{E}} w_{ij} u_i^0 u_j^0 \) is a measure of the affinity within the segment.

Claim: segments with low $\Gamma$ are special and probably correspond to salient regions of the image.

High large affinity within the region and low affinity across the boundary of the region.
Intuitively, SWA coarsens the grid so that segments with low salience are preserved.

6. Define a segment $S^1$ on $\Lambda^1$ by $U^1$

$$U^1_k = 1, \quad \text{if } k \leq S^1$$

$$= 0, \quad \text{otherwise}.$$ 

This has salience at level 1

$$\Gamma^1 [u^1] = \sum_{k, \ell \in \Lambda^1} w_{k,\ell} (u^1_k - u^1_\ell)^2$$

$$\sum_{k, \ell \in \Lambda^1} w_{k,\ell} u^1_k u^1_\ell.$$ 

It corresponds to a segment $S^0$ on $\Lambda^0$ defined by

$U^0_i$ given by $u^0_i = \sum_{k \in \Lambda^1} p_{ik} u^1_k, \quad i \in \Lambda^0$

with saliency

$$\Gamma^0 [u^0] = \sum_{i, j \in \Lambda^0} w_{ij} (u^0_i - u^0_j)^2$$

$$\sum_{i, j \in \Lambda^0} w_{ij} u^0_i u^0_j.$$ 

Claim: if $\Gamma^0$ has low salience, then

1) the $u^0_i$ are close to binary values,

2) $\Gamma^1 [u^1] \approx \Gamma^0 [u^0]$

see papers by Sharon et al.
This approach can be modified to include additional cues.

Main Modification - modify the weights to include texture features.

\[ \Xi \] a node at level \( l \) corresponds to many pixels at the image level.

Compute statistics on these image pixels, similarly for a neighboring node \( \nu \) at level \( l \).

\[ w_{\mu
u} \rightarrow w_{\mu
u} e^{-k|G_{\mu}-G_{\nu}|} \]

where \( |G_{\mu}-G_{\nu}| \) is a measure of the difference of these statistics for nodes \( \mu \neq \nu \) at level \( l \).

(Note: these texture statistics must take into account the fact that pixels are related by interpolation weights).
The saliency measure was proposed by Shi & Malik (normalized cuts). They did not use the hierarchy, or texture statistics.

Connect SWA differs slightly from paper to paper. The details are somewhat arbitrary—some matter, some do not.

The output of SWA is the set of nodes in the hierarchy, together with their saliences. A separate process is required to determine which of the nodes is really salient and should correspond to desirable segmented regions of the image.

But true segments of images may not exist at any level of the hierarchy.

The central region will probably be found as a node in the SWA hierarchy.

But the background will not.