Multimedia Systems and Applications

Auto-cut for web images

James Wang

Based on ACM Multimedia 2009 paper and poster by:

Chen et al., Auto-cut for web images, in Proceedings of ACM Multimedia 2009, pp. 529-532, October 19 - 24, 2009

Problem and Existing Approaches

- **Problem:**
  - Interactive segmentation, whose task is to extract objects from an image using as little interaction as possible, is of great importance in many applications such as photo and movie editing.
- **Approaches:**
  - Boundary-based:
    - requires the user to draw along the object boundary and then adjusts the curve to snap to the real boundary piecewise.
  - Region-based:
    - requires the user to specify two small sets of pixels belonging to the foreground and the background.
    - It is more convenient.

2

Problem and Idea

- **Problems:**
  - *generate ROI cutout automatically for further processing such as image editing, classification and information retrieval.*
  - *Traditional semi-supervised segmentation method requires strong prior knowledge of boundary label.*
- **Idea**
  - *A global cost function is proposed to combine weak prior knowledge with pixel-level feature.*
  - *Compute fuzzy matting components as building blocks to construct semantically meaningful mattes.*
  - *Mattes are hierarchically clustered and ranked by central preference.*

3

Classical Semi-supervised image-cut model

- **The classical semi-supervised learning framework:**

$$f^* = argmin \sum_{k=1}^{n} V(x_i, y_i, f) + \gamma_A ||f||_K^2 + \gamma_I ||f||_I^2$$

$V(x_i, y_i, f)$ **: prior knowledge (such as labeled pixels);**

$||f||_k^2$ **: complexity of the classifier (smooth regulation);**

$||f||_I^2$ **: the intrinsic data distribution (color model).**

- **Intrinsic distribution of data explains the relationship between labels and data affinity.**

4

Intrinsic distribution of data

- **Levin's color-linearity model:**
  - Assume in each local window w:

$$I_i = f_i F_i + (1 - f_i) B_i$$

F_i and B_i are foreground and background colors.

$$f_i \approx aI_i + b, \forall i \in w$$

$$a = \frac{1}{F-B}, b = -\frac{B}{F-B}$$

$$||f||_I^2 = \sum_{j \in I} \left(\sum_{i \in w_i} (f_i - a_j x_j - b_j)^2 + \epsilon a_j^2 \right)$$

The first term reflects intrinsic distribution of pixel colors, while the second term is a regularization term that prefers smooth label in each local window.

- **More details see:**

A. Levin, Rav-Acha, and Lischinski. Spectral matting. IEEE Proceedings on Computer Vision and Pattern Recognition, 2007.

5

Use of Prior Knowledge

- **Levin's color linearity model is very efficient on many images, but it is vulnerable to textured background. Strong edges in background leads to over fractional segmentation.**
- **Use prior knowledge to overcome the problems in Levin's model:**
  - a salient prior to highlight the high contrast region;
  - strokes labeled by users.
- **Hypothesis of this paper:**
  - A "good" image taken by a skillful photographer tends to have its ROI near the center.
  - statistics on about 5,000 web images from various sources such as Google and Flickr reveals that more than 82.2% images have their ROI located at least near the center.

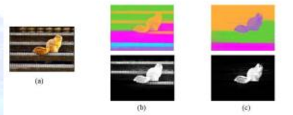
6

A Weak prior on boundary labels

- **assuming that the boundary pixels of an image are less likely to have too fluctuated labels. This rather weak prior reflects central preference.**

$$V(x_i, y_i, f) = \sum_{j \in \partial I} \left(\sum_{i \in w_j} (f_i - \tilde{a}_j x_j - \tilde{b}_j)^2 + \epsilon \tilde{a}_j^2 \right)$$

- **Comparison:**



- (c) is much better than (b).

Cost Function Optimization

- **Foreground/ background label cannot be computed directly because no user interaction is available.**
- **Minimize the quadratic cost function** $J(f) = f^T L f$ **to get semantic components for further processing and integration, where L(I, j) is defined as:**

$$\sum_{q | (i,j) \in w_q} \left(\delta - \frac{1}{|w_q|} \left(1 + (I_i - \mu_q)^T (\Sigma_q + \frac{\epsilon^*}{|w_q|} I_3)^{-1} (I_i - \mu_q) \right) \right)$$

- **Assume that the matting components desired lie in the space spanned by smallest eigenvectors:**

$$E(v_1, v_2, ..., v_m) \text{ of } L$$

- **Compute a rotation matrix** ***R so that ER is sparse:***

$$\sum |f_i^k|^\gamma + |1 - f_i^k|^\gamma \qquad s.t. \sum f_k = 1$$

$$f_k = ER^k$$

Block Integration

- **Idea:**
  - hierarchically cluster semantic components to meaningful parts, and then we rank each of them.
- **Hierarchical Grouping:**
  - use segmentation tree to hierarchically group matting components before labeling.
  - On each level, nodes with very 'similar' statistical properties are combined together as a new component.
  - regional statistics of color to compute the similarity of different components:

$$d(z_i, z_j) = \sum_k EMD_{l1}(h_k(z_i), h_k(z_j))$$

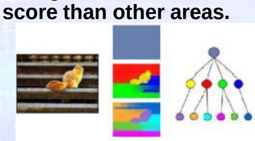
Block Integration (cont.)

- **Central preference ranking:**
  - Rank each component by the normalized central preference score:

$$H(z') = \frac{\sum_{i \in I} I(i \in z'_j)(1 - e^{-\frac{d^2}{\sigma^2}})}{\sum_{i \in I} I(i \in z'_j)}$$

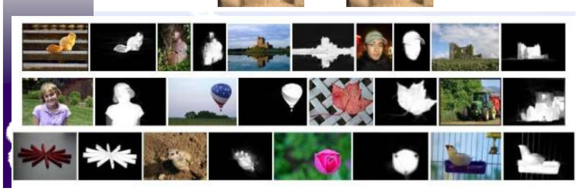
where *σ reflects the degree of our central preference*.

- **In most images, the ROI tends to have much higher score than other areas.**



Experimental Results

- **The results are quite satisfactory.**



Conclusion and Future Work

- **A novel automatic segmentation algorithm has been proposed.**
  - Use central preference as the weak prior knowledge.
  - Develop a 3-step segmentation procedure :
    - compute semantically meaningful components through spectral clustering;
    - group the oversegmented components based on their regional statistics;
    - rank the segments by central preference.
- **The nice performance on a large benchmark data.**
- **In the future, incorporate more information such as the face detector and saliency detector in ROI selection.**