Region-Based Multiscale Spatial-Temporal Saliency for Object Detection from Video

Trung-Nghia LE, Akihiro SUGIMOTO

Motivation

Detecting salient objects in videos for multimedia and computer vision applications.

Robot Navigation  Augmented Reality  Video Retarget

Approach

Combining static and dynamic features from the low and middle levels in a segmentation pyramid model.

Utilizing spatial features of each frame and temporal features across frames.

Developing adaptive temporal window to keep temporal consistency across frames.

Proposed Method

Multiscale Segmentation Pyramid


Spatial Saliency Entity Construction

Contrast Calculation

Linear Calculation

Incorporating temporal consistency

Adaptive Temporal Window

Spatial Saliency Entity Construction

Color  Intensity  Orientation

Flow Orientation  Flow Magnitude

Low-level Features

Low-level Feature Map

Middle-level Features

Middle-level Feature Map

Spatial Saliency Entity

Multiscale Fusion


Evaluation and Results

Evaluation on 4 public datasets (Weizmann, 10-Clips, JHMDB, MCL) by comparing with several state-of-arts (LC[1], SAG[2], STS[3], HS[4], SO[5]).

1) Weizmann  2) 10-Clips  3) JHMDB  4) MCL
