

A Computational Model of Saliency Depletion/Recovery Phenomena for the Salient Region Extraction of Videos

July 03, 2007

Media Information Laboratory

NTT Communication Science Laboratories

Nippon Telegraph and Telephone Corporation

Clement Leung ■ Akisato Kimura ■ Tatsuto Takeuchi ■ Kunio Kashino

Overview

- Objective and related works
- Basic Algorithm Structure
- New computational model:
 - Instantaneous Saliency Depletion with gradual Recovery
 - Long term Saliency Depletion with instantaneous Recovery
- Algorithm Evaluation against previous algorithms using eye tracking tests
- Summary

Objective

- We use the strategy of focusing on more relevant regions and suppressing irrelevant regions in videos
 - Reduces the amount of data to be processed
 - Only unique information is retained
- Probable approach:
Computational model is established based on the human visual system
 - Human Vision has a powerful ability to extract important information from a given scenery

Related Works

Itti, Koch & Niebur (1998)

- Still Image Algorithm:
 - Proposed a model for computing saliency from still images
 - Features: Intensity, Color & Orientation
 - **Restricted to still Images**

Itti, Dhavale & Pighin (2003)

- Moving Algorithm:
 - Added onto the previous model flicker and motion features to produce video saliency extraction
 - **Did not take into account the temporal dynamics of the human visual system**

New Computational Model

We have extended the previous algorithms to include two important temporal characteristics:

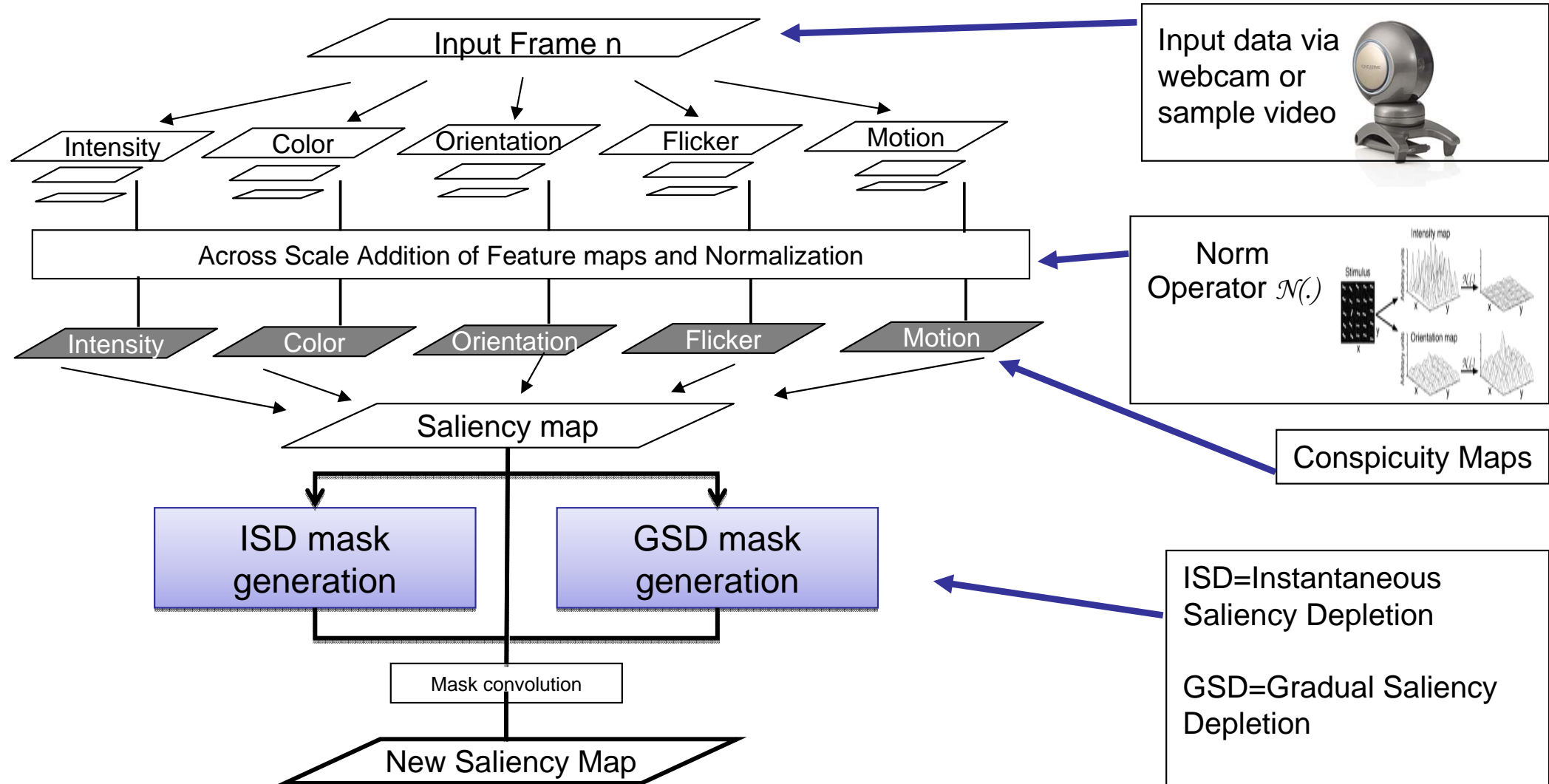
1. Instantaneous Saliency Depletion with Gradual Recovery

- Based on “*Inhibition of Return*” theorem (Posner 1984):
human attention tends to have a delay in realizing salient events around regions previously focused on

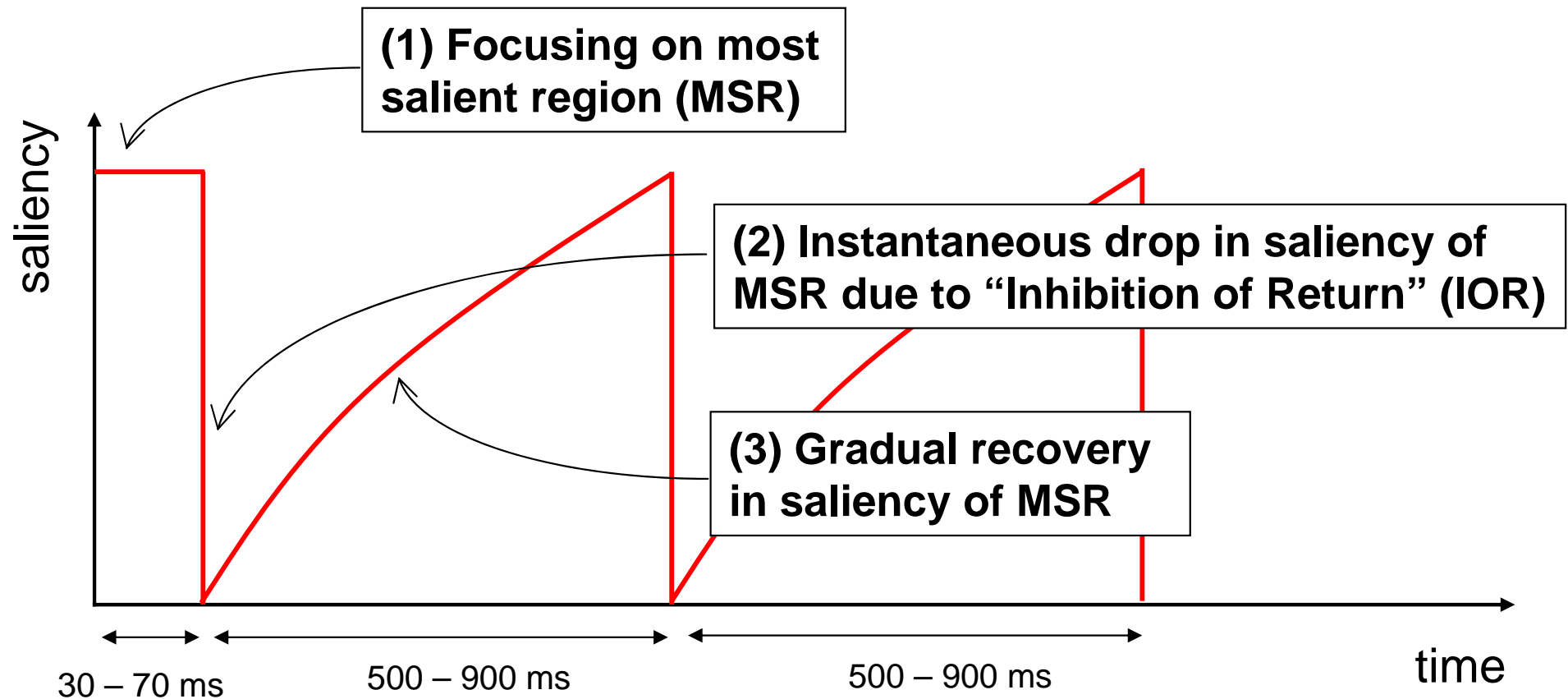
2. Gradual Saliency Depletion with Instantaneous Recovery

- Based on “*Neural Adaptation*” theorem (Hartline 1940):
saliency gradually decreases over time when no surprising events occur in a video

Basic Algorithm Structure



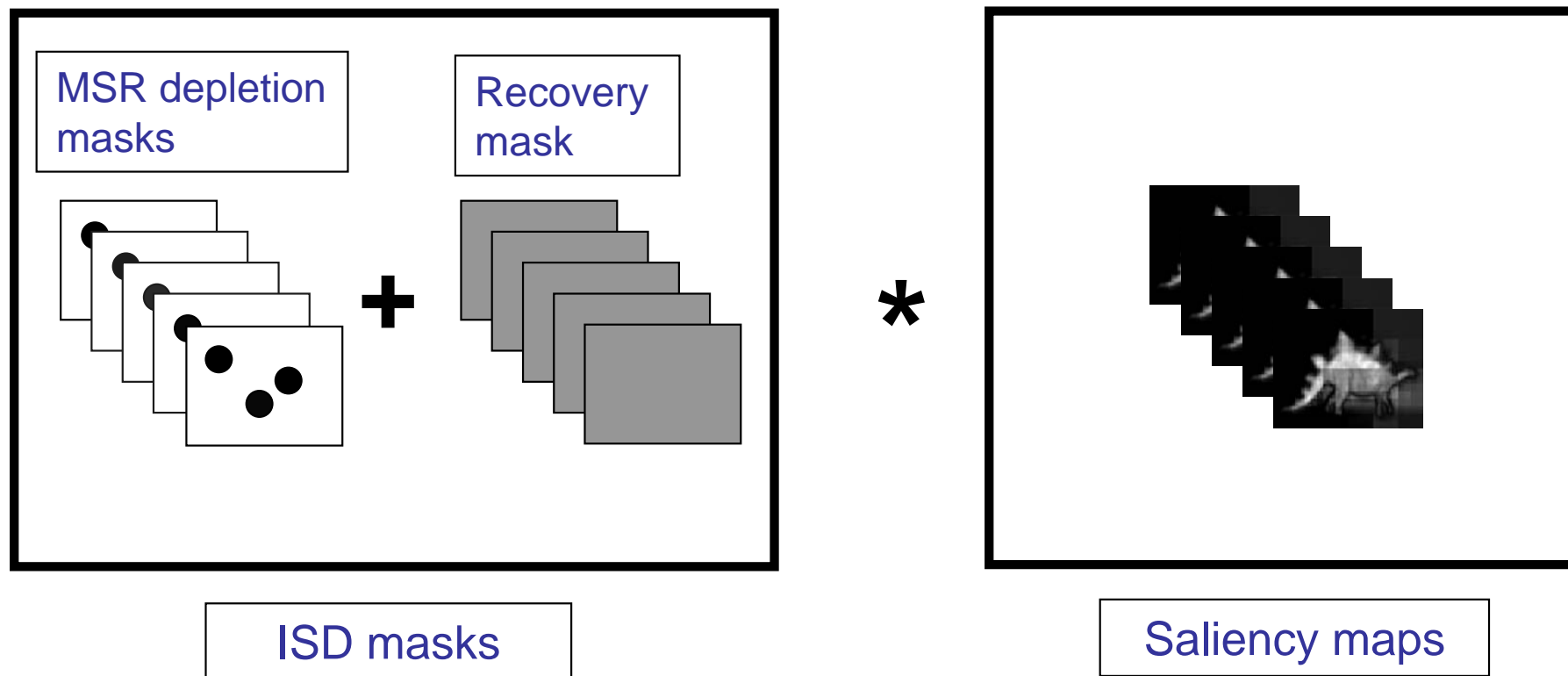
Instantaneous Saliency Depletion: Graphical Interpretation



(Saliency drop and recovery times are based on the IOR theorem)

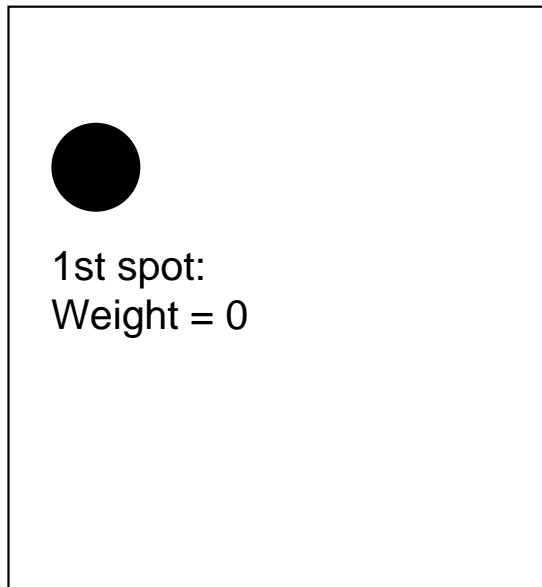
Instantaneous Saliency Depletion: Implementation Strategy

- Instantaneous saliency depletion (ISD) mask is created for each frame
- Multiplied with corresponding saliency Map

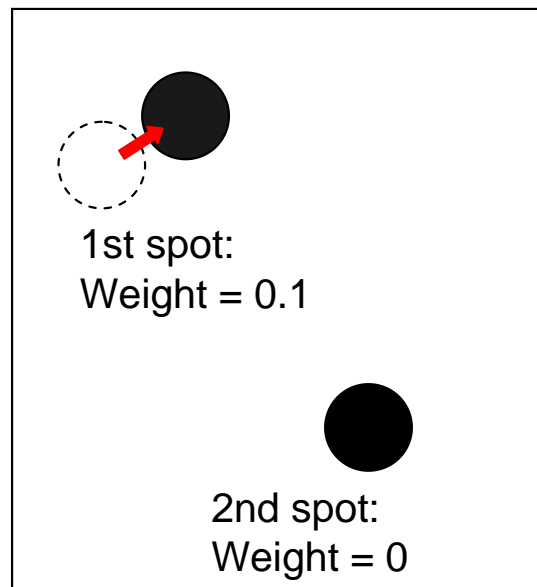


Instantaneous Saliency Depletion: MSR Depletion Masks

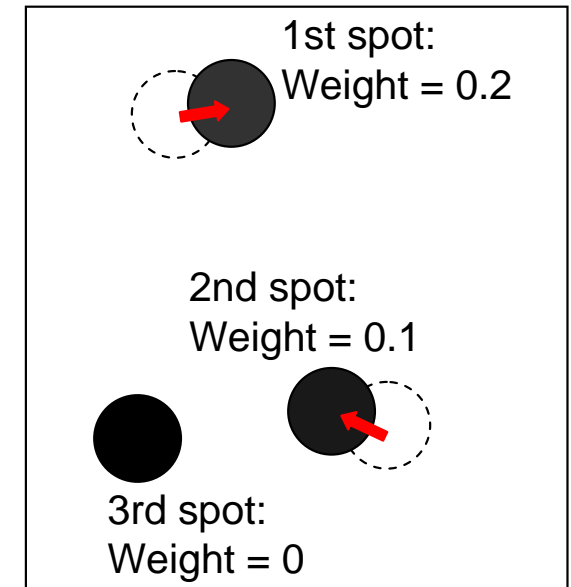
- Saliency depletion regions move accordingly with object(s) in motion.
- New coordinate values of depletion regions are found using X and Y optical flow information.



Frame 1: MSR of frame 0 is blacked out



Frame 2: The MSR of frame 1 is blacked out, while 1st MSR spot starts to recover



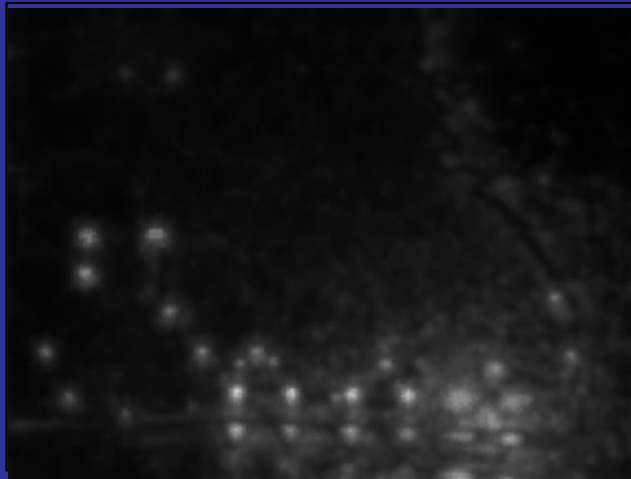
Frame 3: MSR of frame 2 is blacked out while previous two MSR spots recover

Instantaneous Saliency Depletion: Example

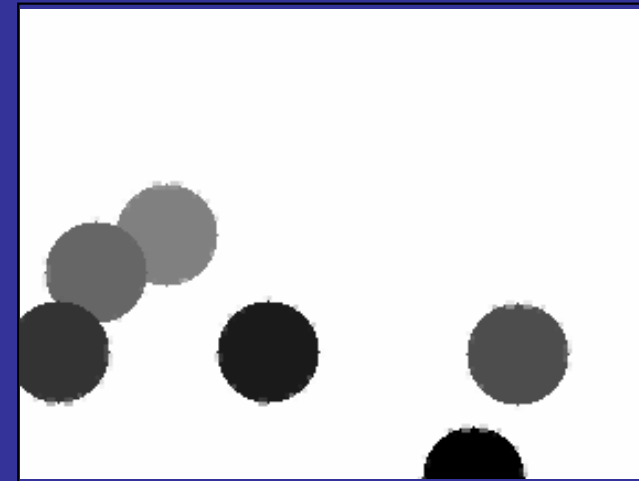


Original Video

Instantaneous Saliency Depletion: Example



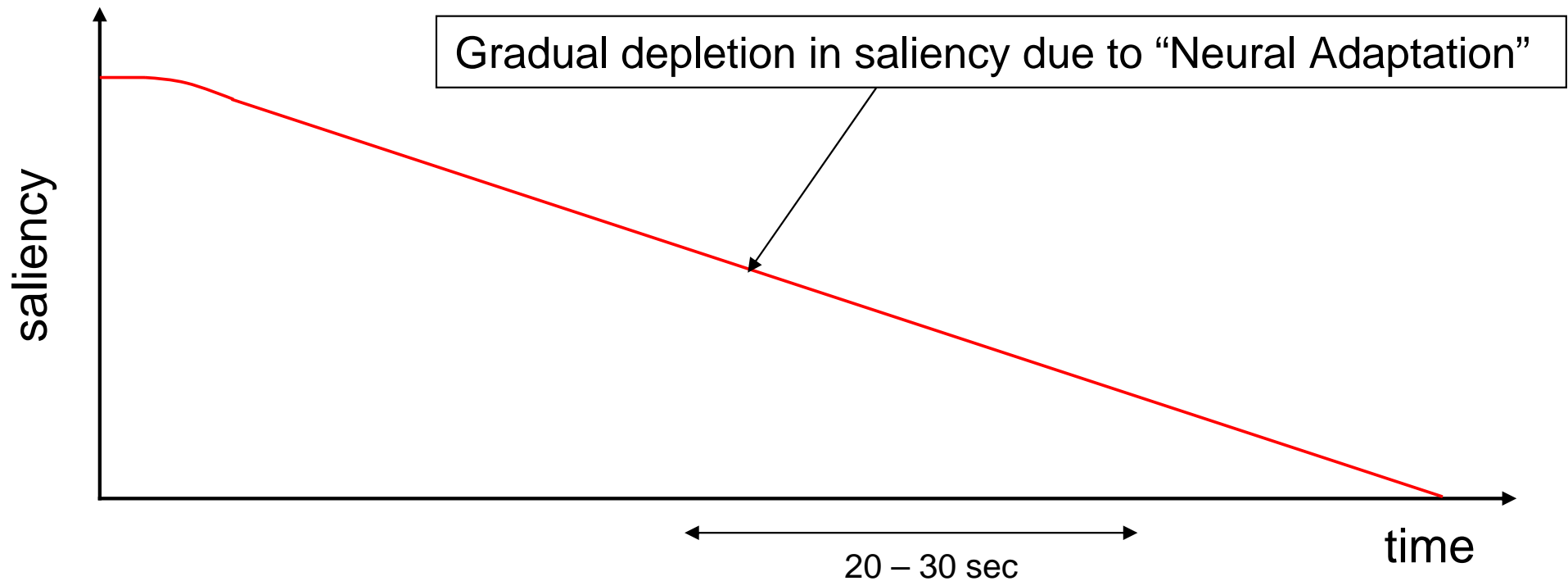
Saliency with instantaneous saliency depletion



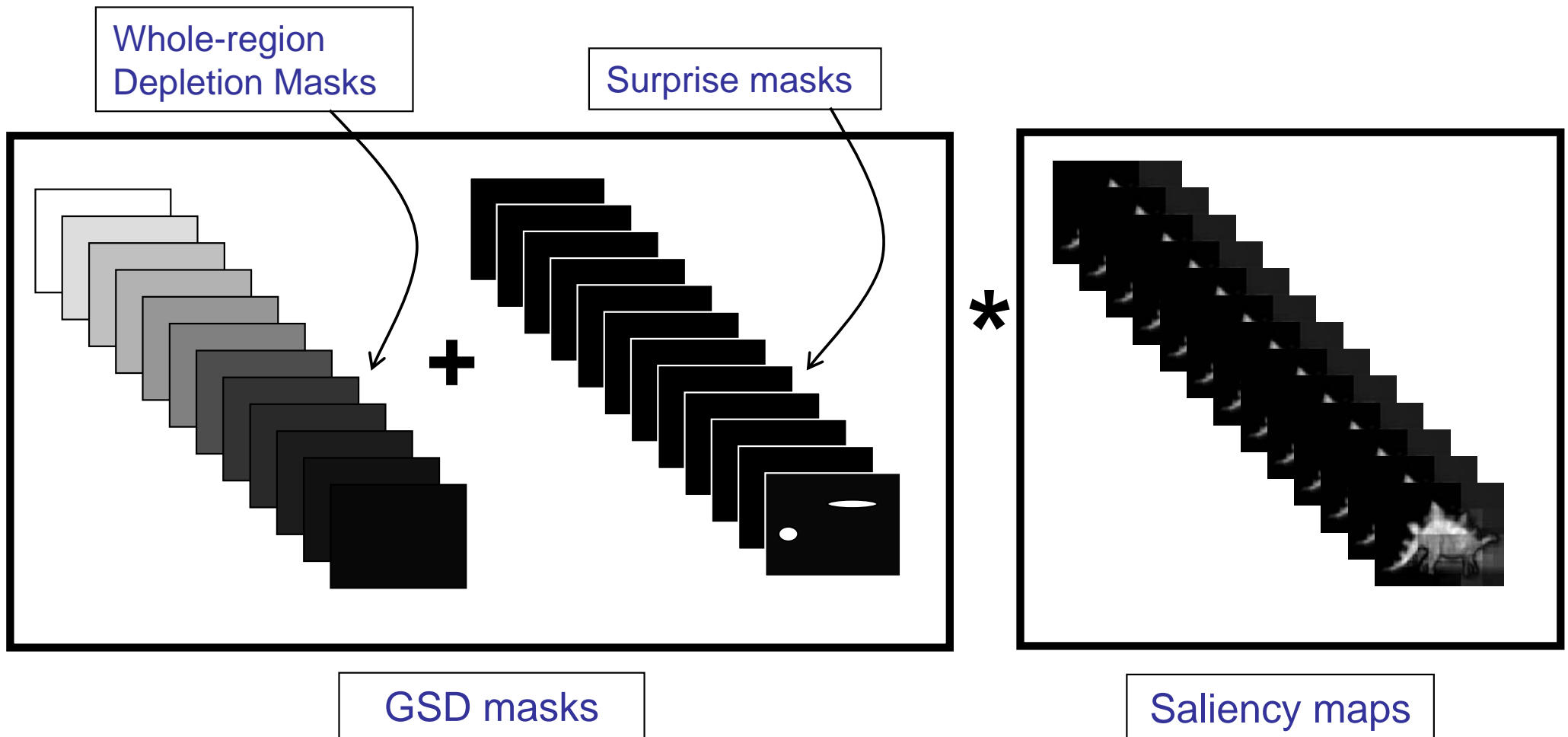
Instantaneous saliency depletion/recovery mask

Gradual Saliency Depletion: Graphical Interpretation

Gradual saliency depletion with instantaneous recovery for cases such as still unchanging videos, constant velocity, constant flickering pattern.

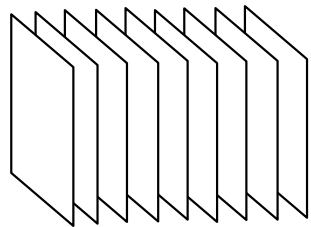


Gradual Saliency Depletion: Implementation Strategy



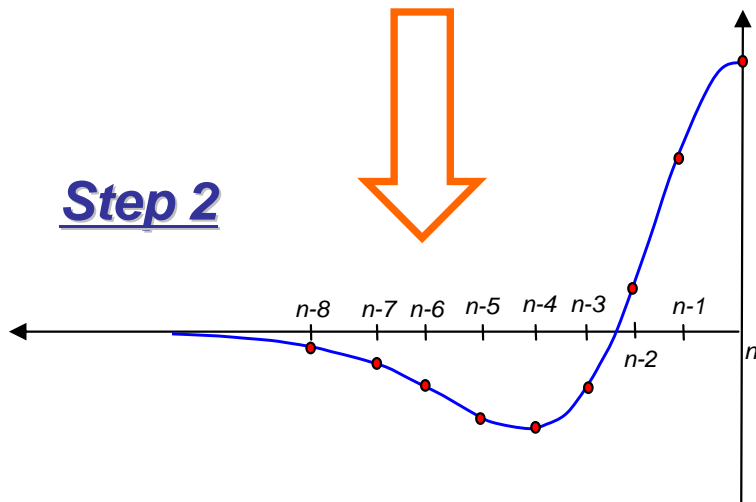
Gradual Saliency Depletion: Surprise Masks

Step 1

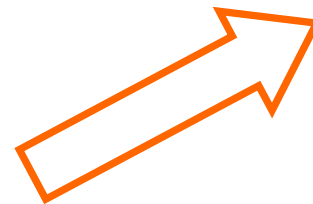


Conspicuity map frames $n-8 \dots n$ for a specific feature (e.g. intensity)

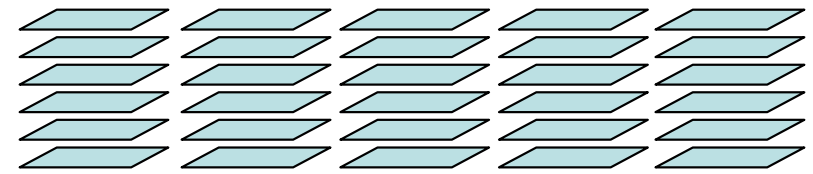
Step 2



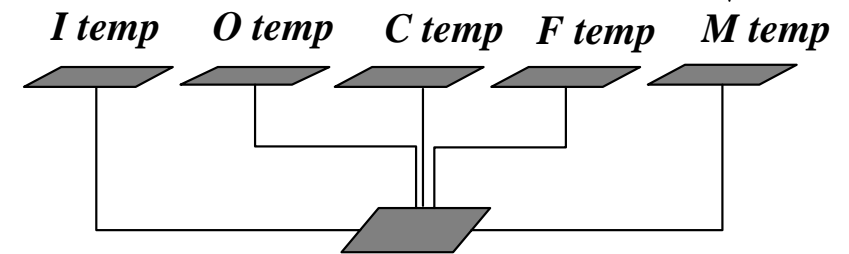
Extract temporal feature maps through difference-of-Gaussian (DoG) filtering on the temporal domain



Temporal Feature maps



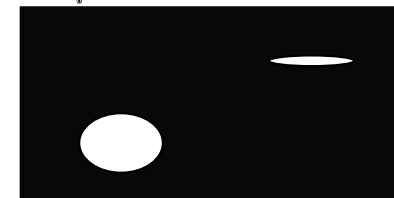
Across scale addition of Temporal Feature maps



Sum of temporal feature map

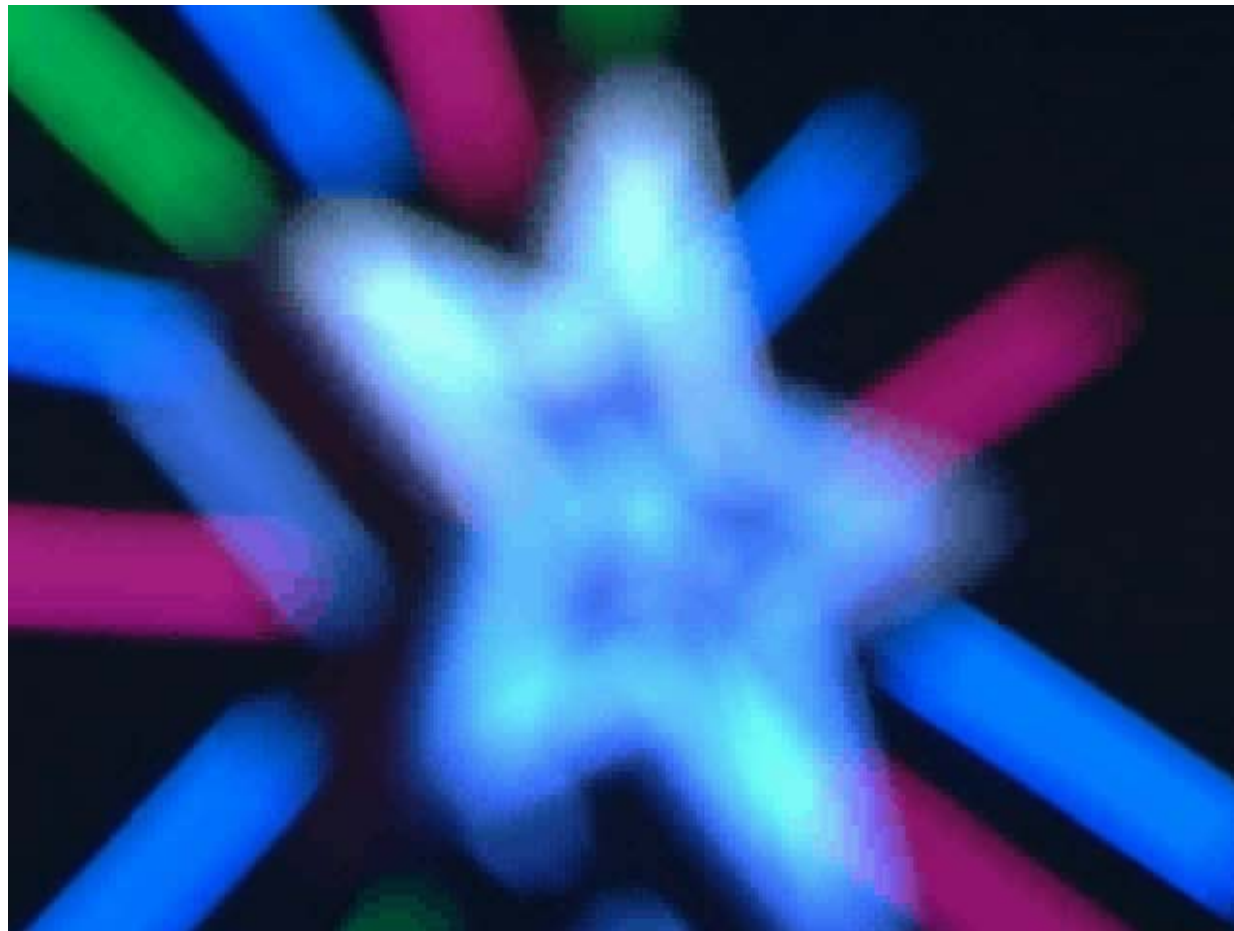
Step 3

Create a surprise mask from temporal feature maps



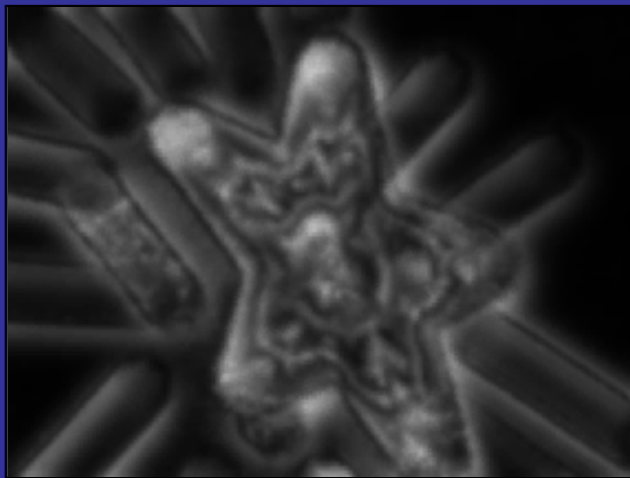
Surprise mask

Gradual Saliency Depletion: Example



Original Video

Gradual Saliency Depletion: Example



Saliency with gradual
saliency depletion



Gradual saliency
depletion/recovery mask

Algorithm Evaluation

Procedures:

- 5 subjects, each view 6 different sample videos while we track their eye movement
- Compare eye tracking test results to saliency videos produced by all three algorithms:
 - a.) Itti's Still Image Algorithm
 - b.) Itti's Moving Algorithm
 - c.) Our Algorithm:
 - Case 1: instantaneous depletion/gradual recovery only
 - Case 2: gradual depletion/instantaneous recovery only
 - Case 3: with both depletion properties included

Algorithm Evaluation

Measures for evaluation:

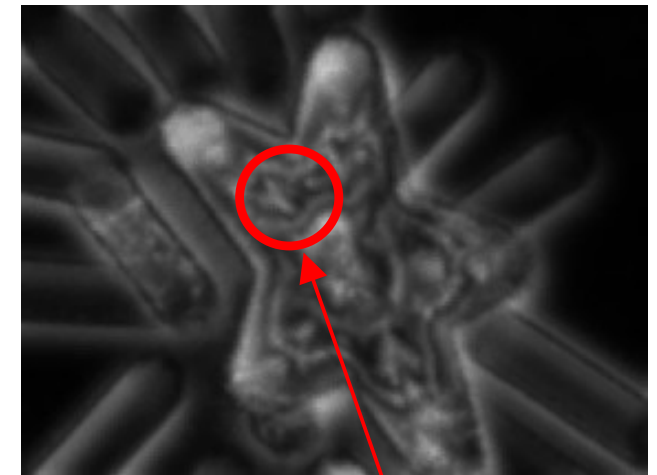
- Located region where the eye is focusing on in each saliency video frame
- Calculated the normalized average pixel value at the region

NETR value =

Avg Pixel value in Eye focusing region

—————
Total pixel sum of frame

Eye focusing region: center=the eye tracking point,
radius=height of frame/9.

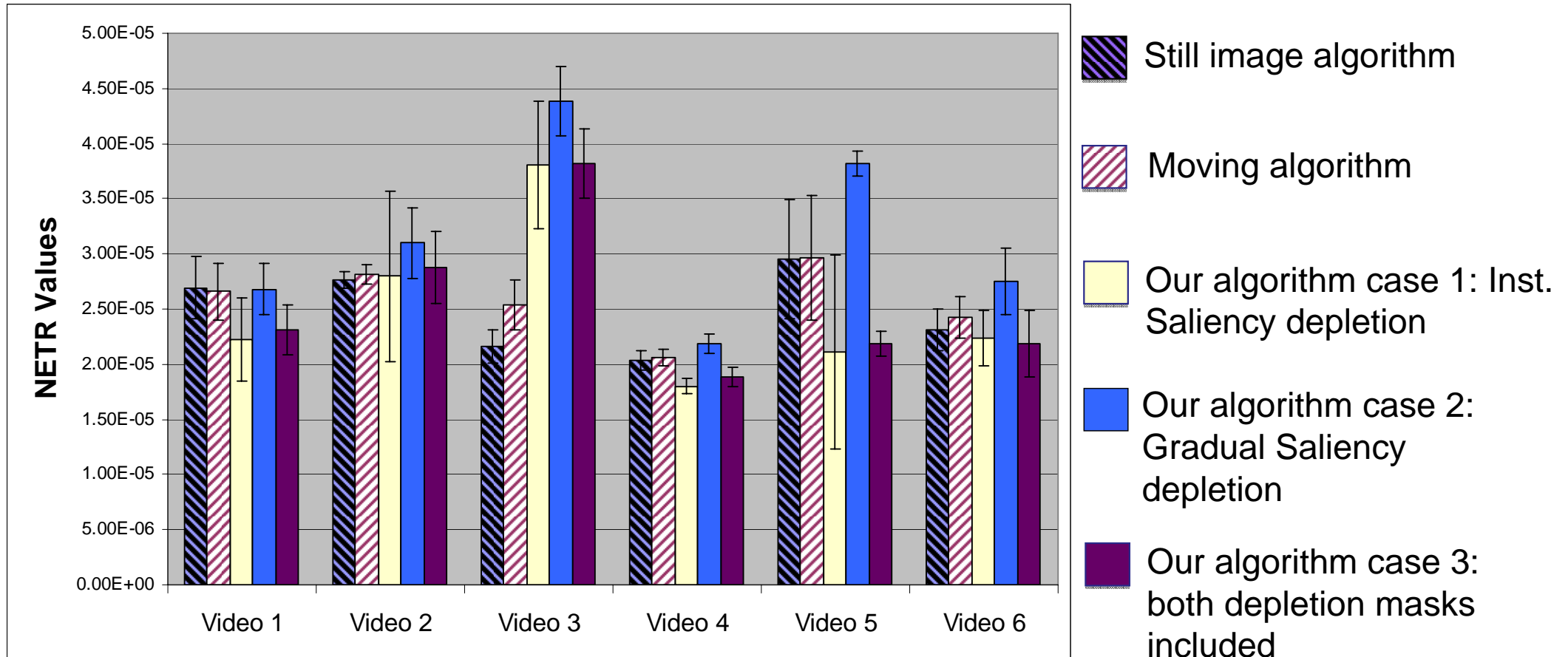


NETR

Best performance:

High Average Normalized Eye tracking region value.

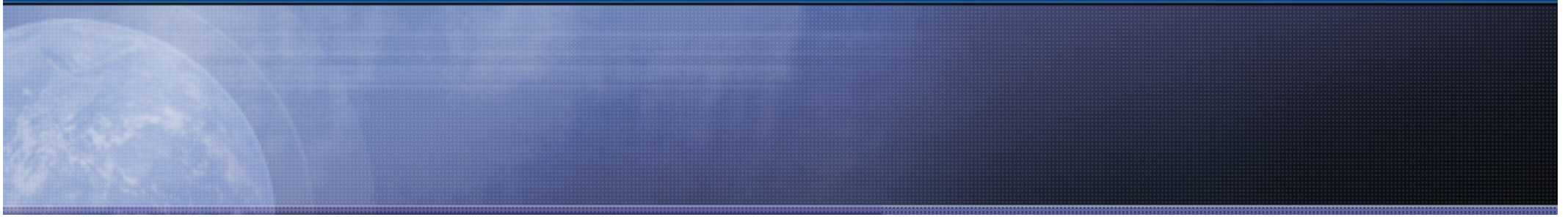
Results



- Overall, our algorithm has approximately the same or substantially better performance than the other two algorithms

Summary

- Algorithm is based on extraction of early visual features to create saliency maps
- **Instantaneous Saliency Depletion:** Based on the “*Inhibition of return*” theorem, attention instantaneously diverts away from an area after being attended to, following by gradual saliency recovery
- **Gradual Saliency Depletion:** Based on “*Neural Adaptation*” theorem, a gradual decrease in saliency over time, with instantaneous saliency recovery of surprising areas
- Test results show that our algorithm performs approximately the same as previous algorithms or substantially better depending on selection of video
- Some demonstration movies will be available at <http://www.brl.ntt.co.jp/people/akisato>



Future Plans

- Do more tests with more subjects using artificial videos
- Conduct further studies to find more precise duration of time for Instantaneous/Gradual saliency depletion and recovery
- Incorporate Color-Intensity relationship
- Introduce High level knowledge Strategies: Learning algorithms to consider people's preferences.

Analysis of Results

Video 3

Video 5

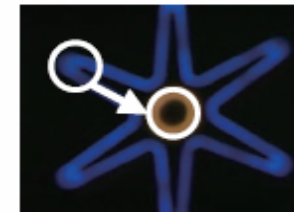
Original Video



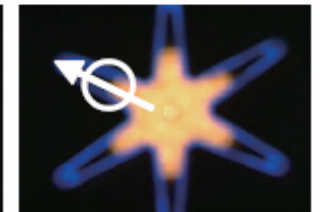
Pedestrian "STOP" sign is lit while background signals are blinking.



The pedestrian sign changes to "WALK".

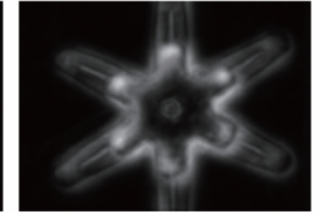
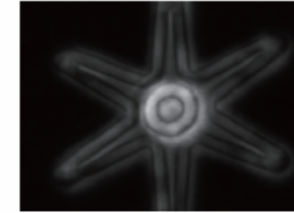


Center of blue star is shining.

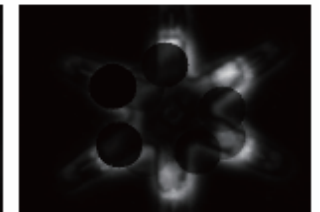
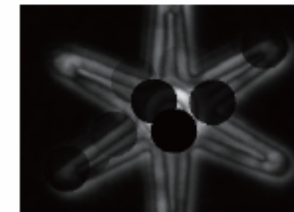


Yellow light radiates in blue star.

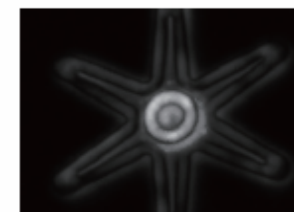
Moving algorithm



Our Algorithm Case 1



Our Algorithm Case 2



Analysis of Results

High Standard Error in our Algorithm's Results for each video due to:

1. High level knowledge
2. Limited number of test subjects
3. In our algorithm: If eye is not at salient region very low values, if eye is on salient region very high values.
4. Previous algorithms: Not as many regions are suppressed, if eye is not on most salient region, there are still other regions that are less salient but will give higher values than in our algorithm.