Object Recognition: From Man to Machine

Gert Kootstra, Artificial Intelligence, University of Groningen, The Netherlands

1. Paying Attention

Study 1: Where Do Humans Pay Attention To?

Introduction

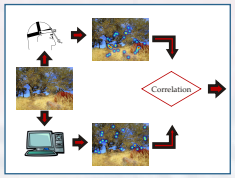
- People view actively: using eye movements to select interesting information.
- Eye movements are top-down (model-driven) and bottom-up (stimulus-driven) controlled.
- Question: Can we predict eye movements with BU saliency models? And how?

Methods

- Different saliency models
  - Saliency model of Itti et al. (1998)
  - SIFT keypoint detection (Lowe, 2004)
  - Symmetry models (Heidemann, 2004; Reisfeld et al., 1995)
- Eye tracking experiment
  - 43 participants
  - Free-viewing
  - 4 categories of 10 images

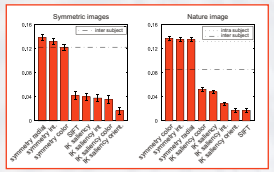
Measure

- Correlation between saliency models and human fixation density maps.



Results

- High correlation symmetry models.
- Also on non-symmetrical images.
- Saliency model and SIFT correlate less, but still significantly.



Discussion

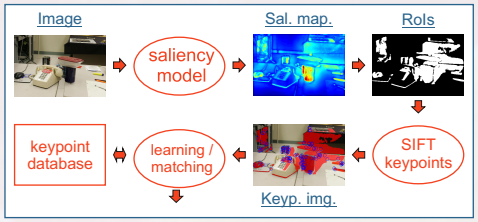
- Symmetry seems to attract human visual attention.
- Future work: test symmetry for computer vision

Study 2: Using Attention Models for 3D Object Recognition

Introduction

- Use saliency models to select regions of interest (RoI) in images.
- Gather SIFT keypoints only from RoI's.

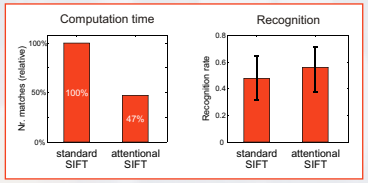
"Attentional SIFT" model



Experiments

- Image dataset (Kushal & Ponce, 2006)
  - 9 objects from 7-12 different viewpoints.
  - 80 images with objects in cluttered scenes.

Results



- Attentional SIFT uses far less keypoints, which improves computation time.
- Moreover, recognition performance, measured by nr of correct matched keypoints, improves slightly (though not significant).

Discussion

- Using attentional (saliency) models improves 3D object recognition with SIFT.
- Future work: test on larger datasets.

2. Exploring the Objects

Study 3: Active Vision for 3D Object Recognition in Real-World

Introduction

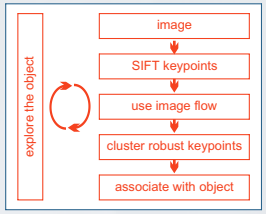
- Difficulties for object recognition in the real-world:
  - Object constancy problem (see fig.)
  - Objects are positioned in highly cluttered environments

- Solution in nature: exploration (see fig.)

Exploration to view objects from multiple viewpoints and to segment them from background

Model

- Exploration: construct 3D model by moving around the object
- Representation of object by collection of SIFT keypoints from different views
- Active vision to segment object from background using optical flow:
  - Object points move consistent with each other
  - Background points have different motion vectors.

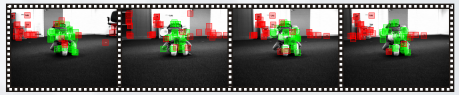


Keypoint clustering

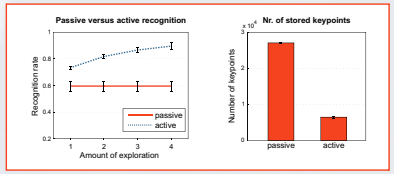
- Standard SIFT results in many keypoints. We use a clustering method to cluster keypoints.
- Smaller keypoint database improves computation time for recognition.
- Clustering method:
  - Growing When Required (GWR) network (Marsland et al. 2002)
  - Based upon Kohonen SOM, but adjusts number of nodes (=clusters)

Results

- Successful segmentation from background



- Large improvement in learning and recognizing in cluttered environments using active vision, while using less keypoints



- The GWR-SIFT reduces the amount of keypoints with 36%, while resulting in a better performance than standard SIFT with 36% keypoints

Discussion

- Exploring objects makes object recognition in the real-world possible
- GWR reduces nr of keypoints. More research to increase recognition

Heidemann, G. (2004) Focus-of-attention from local color symmetry. *IEEE Transaction of Pattern Analysis and Machine Intelligence*, 26(7): 817- 830.
Itti, L., Koch, C,& Niebur, E. (1998) A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11): 1254-1259.
Kushal, A. &Ponce, J. (2006) Modeling 3D objects from stereo views and recognizing them in photographs. *Proc. European Conference on Computer Vision*, 2006.
Lowe, D. (2004) Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2): 91-110.
Marsland, S., Shapiro, J. & Nehmzow U. (2002) A self-organising network that grows when required. *Neural Networks*, 15: 1041-1058
Reisfeld, D., Wolfson, H., Yeshurun, Y. (1995). Context-free attentional operators: the generalized symmetry transform. *International Journal of Computer Vision*, 14(2): 199-130.