Data-driven Approach to Dynamic Visual Attention modelling

Dubravko Culibrk[1], Srdjan Sladojevic[1], Nicolas Riche[2], Matei Mancas[2] and Vladimir Crnojevic[1]

[1]University of Novi Sad, Serbia

[2]University of Mons, Belgium

Motivation

- Eye-tracking data usually used for verification of the various models of visual attention (saliency).
- Could potentially also be used in conjunction with machine learning techniques to learn the models of attention.
- This venue is just beginning to be explored.

UMONS
Université de Mons

Problems learning from eye-tracking data: too much data

- The visual stimulus is an image or a video in case of dynamic attention and the goal of a attention model is to predict whether a pixel of an image or a frame is of interest or not.
- For a standard resolution video (720 x 540), one has to deal with environ 9 million points per second.
- State of the art machine learning algorithms are unable to handle such large amounts of data.

UMONS
Université de Mons

Problems learning from eye-tracking data: precision/variance

- Eye-tracking data is usually fairly imprecise due to the acquisition process and unsuitable for determining the level of interest on a per-pixel level.
- The eye-tracks of subjects viewing the same scene vary to a great extent.
- Rather than aggregate the data beforehand, the machine learning algorithm should be able to generalize across the different subjects.

UMONS
Université de Mons

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА НОВИ САД

Our approach

- Novel data sampling methodology, which allows us to create a representative training and testing data set from a database of 24 videos (eye-tracks of 13 users).
- Conventional dynamic saliency approach to used to extract basic features.
- Conventional machine learning algorithm (decision tree) used to learn how to classify between interesting and other pixels.
- The resulting visual attention model verified using standard methodology.

UMONS
Université de Mons

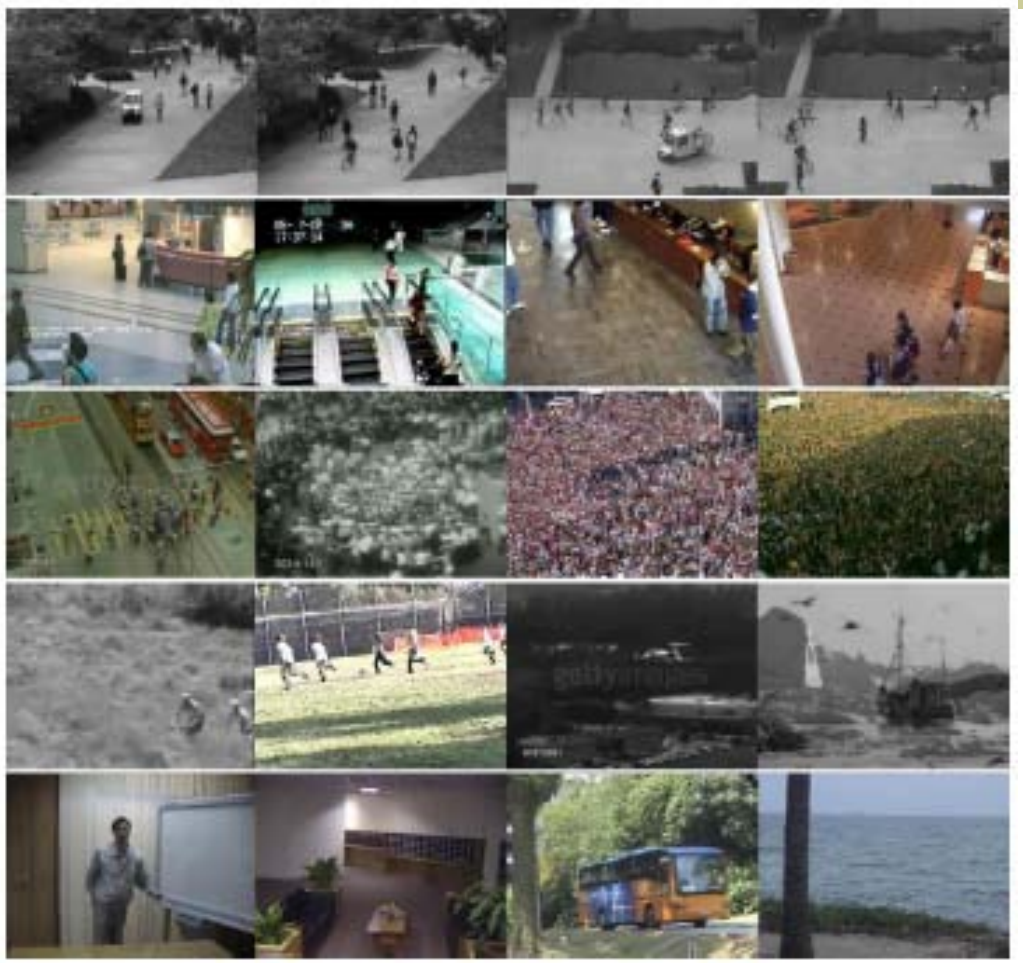
Eye tracking data

- ASCMN database, 24 videos and eye-tracks collected using FaceLab.

Table 1. The 5 classes of videos contained into the ASCMN database

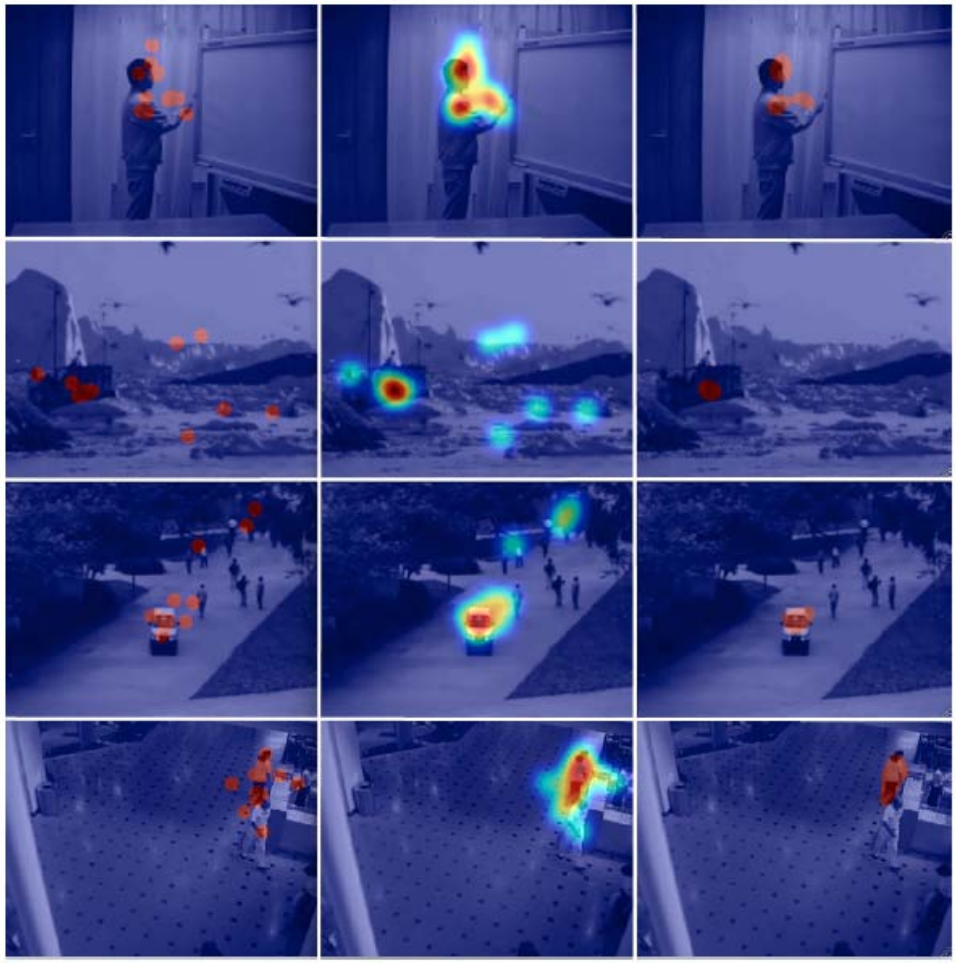
Video classes	Description	Videos Nb.
1) Containing abnormal motion (**ABNORMAL**)	Some moving blobs have different speed or direction compared to the main stream: Figure 2 line 1.	2, 4, 16, 18, 20
2) Video surveillance style (**SURVEILLANCE**)	Classical surveillance camera with no special motion event: Figure 2 line 2.	1, 3, 5, 9
3) Crowd motion (**CROWD**)	Motion of more or less dense crowds: Figure 2 line 3.	8, 10, 12, 14, 21
4) Videos with moving camera (**MOVING**)	Videos taken with a moving camera: Figure 2 line 4.	6, 19, 22, 24
5) Motion noise with sudden salient motion (**NOISE**)	No motion during several seconds followed by sudden important motion: Figure 2 line 5.	7, 11, 13, 15, 17, 23

ASCMN Sample Frames



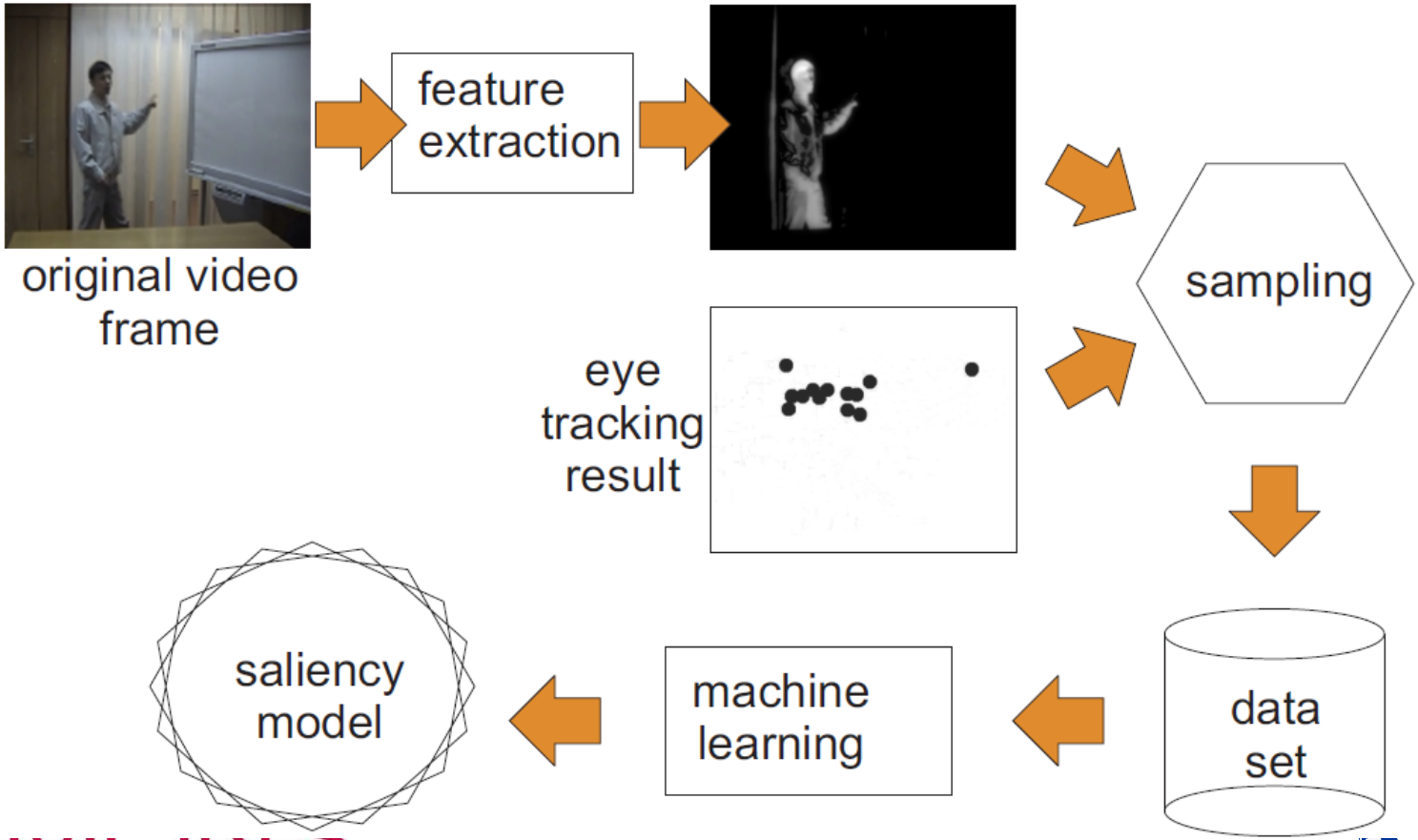
UMONS
Université de Mons

Aggregated Eye-tracks = Saliency



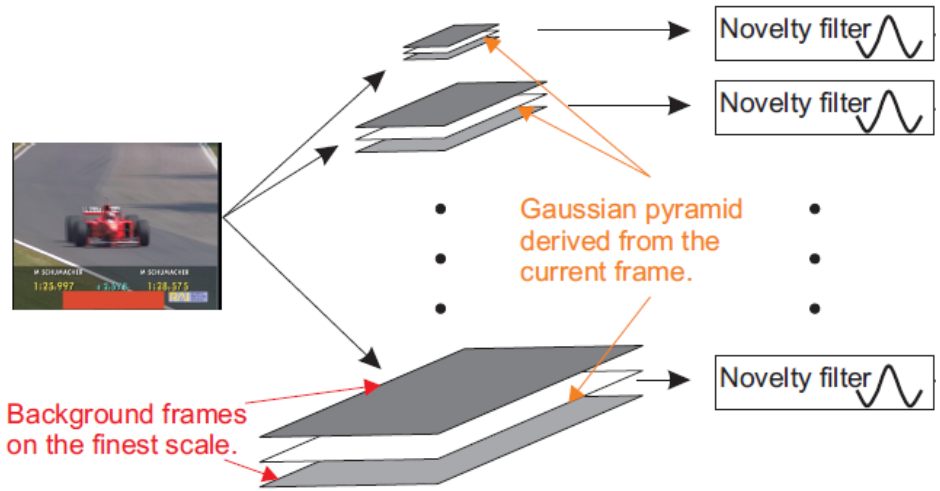
UMONS
Université de Mons

Data-driven saliency modelling



Basic features

- Same as used by a conventional dynamic saliency model - change detection features across a Laplacian pyramid.



UMONS
Université de Mons

Data sampling

- The output of the eye-tracker is represented as a set of circular regions.
- We randomly sample pixels using both the eye-tracking binary mask and the set of extracted saliency features. Stratified sampling (same number of pixels of each class).
- Class: 0-nonsalient,1-salient.

UMONS
Université de Mons

Constrained data sampling

- Due to the rather large diameter of regions marked in the eye-tracking maps, the values of saliency features extracted for large numbers of pixels within these regions were equal to zero. S
- Since such data samples would not benefit the learning algorithm, these points have been discarded.
- Sampling methodology modified to extract salient point samples, only when at least one feature has a value larger than zero.

UMONS
Université de Mons

Train/test dataset

- 10 salient and 10 non-salient points from each frame of the videos in the ASCMN database.
- First 10 frames were discarded to ensure that the saliency features are stable.
- The final dataset contains environ 198,000 samples.

UMONS
Université de Mons

Machine learning

- Used a grafted decision tree algorithm available within the data-mining and machine learning tool Wakaito Environment for Knowledge Analysis (WEKA).
- The final classifier fairly simple and easy to implement.
- WEKA is able to generate Java code for the trained decision tree.

UMONS
Université de Mons

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА НОВИ САД

Results – sample frames

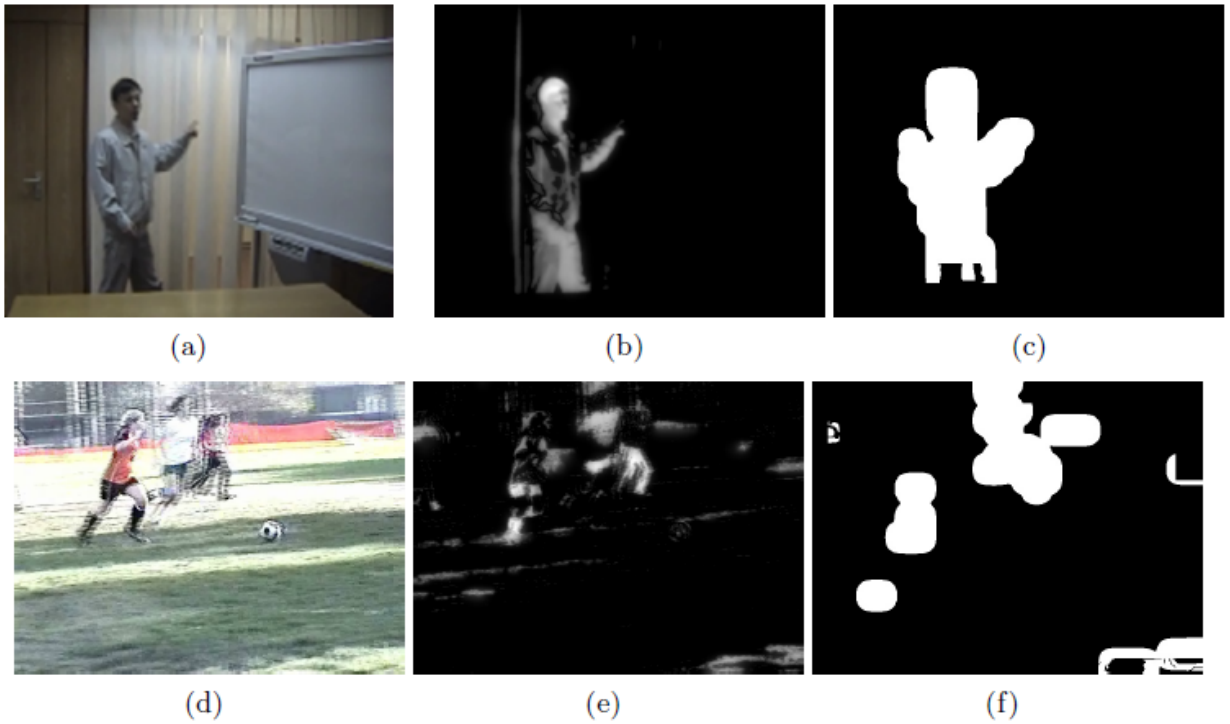


Figure 5. Sample frames for sequences 7 and 19 and the proposed approach: a,d-original frames, b,e-features (sum), c,f-detection result

Results - quantitative

- Overall accuracy of the proposed approach achieved is 84%.
- The salient points were accurately detected in 95% of the cases.
- The non-salient points were detected accurately in just 73% of the cases.

UMONS
Université de Mons

Comparison

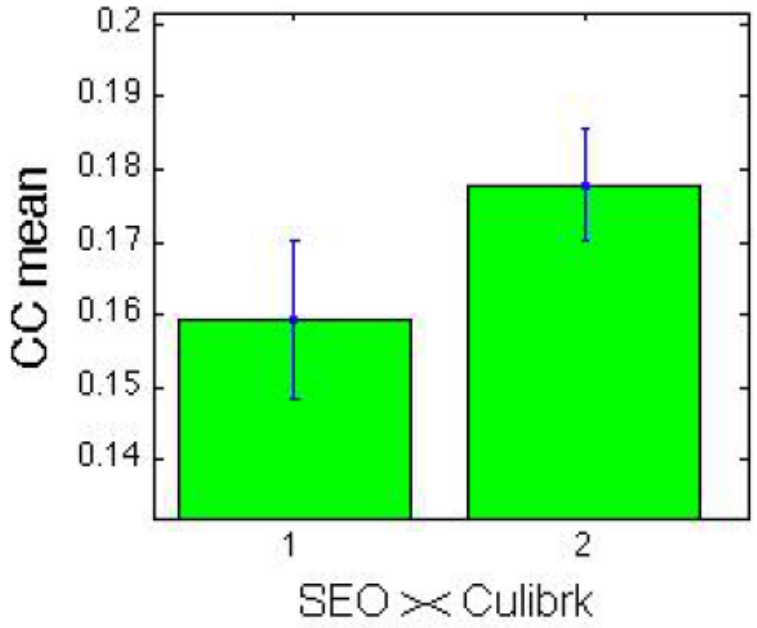
- Compared against state-of-the-art bottom-up saliency model of Seo *et al.*

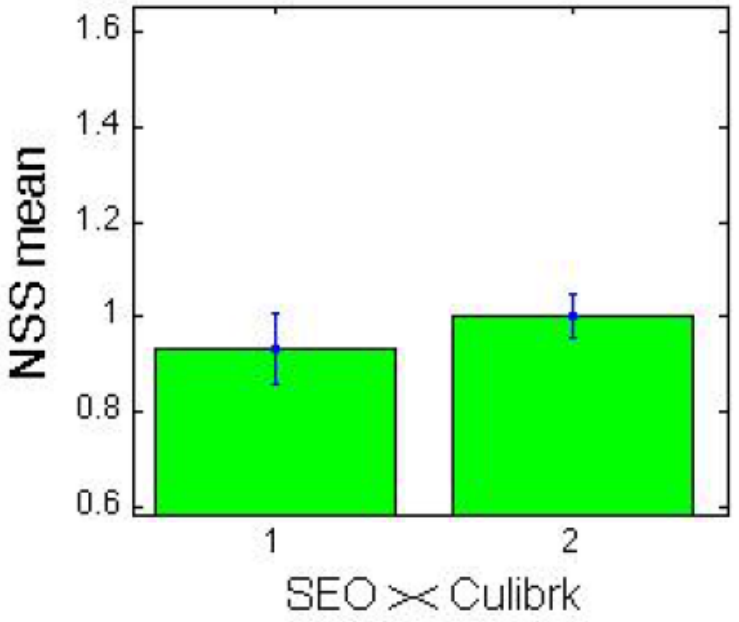
- Measures used:
  - Normalized Scanpath Saliency (NSS) - the average of the response values at human eye positions in a model's saliency map.
  - Correlation Coefficients (CC)

UMONS
Université de Mons

ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА НОВИ САД

Comparison results





UMONS
Université de Mons

Conclusion

- Novel approach dynamic saliency detection presented.
- Framework proposed for learning the saliency model directly from the eye-tracking data.
- Arbitrary saliency-related features can be used for this purpose and diverse machine learning algorithms can, subsequently, be trained to achieve saliency detection.
- State-of-the-art comparable performance achieved using simple features and learner (decision tree).

UMONS
Université de Mons

Thank you!