Towards Space-Time Semantics In Two Frames

Karla Brkić[1], Axel Pinz[2], Zoran Kalafatić[1], Siniša Šegvić[1]

[1]Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia

[2]Institute of Electrical Measurement and Measurement Signal Processing, Graz University of Technology, Austria

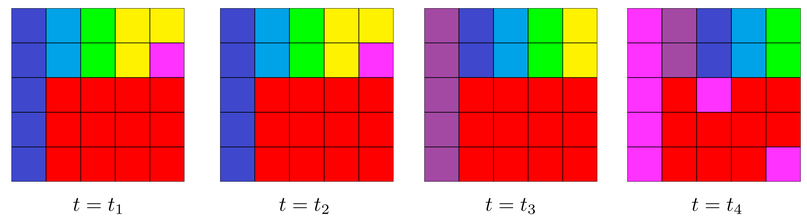
key idea

What can we discover about local scene structure using only two frames?
Assume that a region of interest is divided into a regular grid of patches and correspondences exist between frames. We propose a semantically meaningful descriptor which assigns weights to the following four hypotheses:

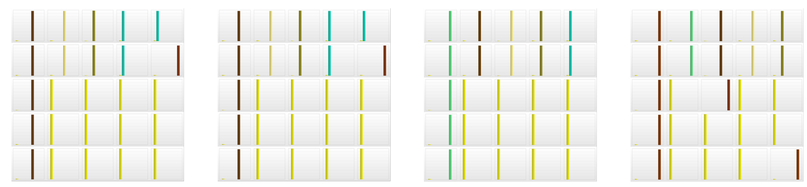
- **C**onstant, for a patch whose appearance does not change between the observed two frames
- **O**uter, for a patch having been changed from the outside of the region of interest
- **I**nner, for a patch having been changed "by itself", i.e without outer influence
- **N**eighbor, for a patch having been changed by a neighboring patch in the region of interest

algorithm

Example synthetic sequence:



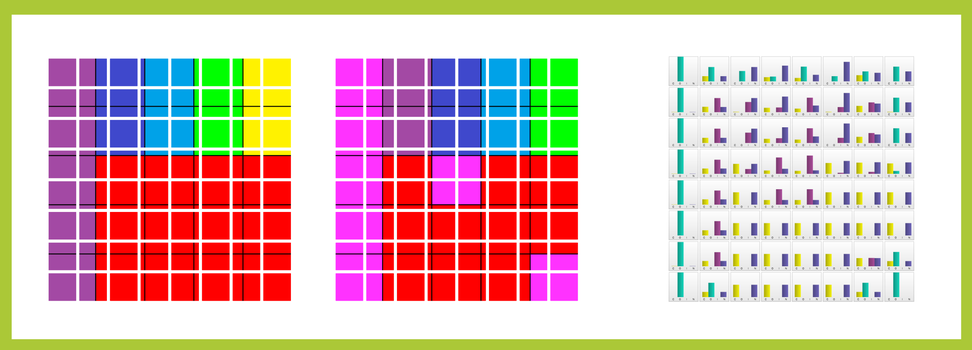
Hue histograms of individual frames:



Calculated COIN descriptors:

1. Find the difference between the patch histogram in the current frame and the patch histogram in the previous frame.
2. Find the most similar 4-neighbor, so the mass of the intersection of the observed patch and the neighbor is maximal.
3. The joint weight of the inner and the outer hypothesis is calculated as the change which can neither be explained by neighbor nor by constant hypotheses, using values from steps 1 and 2. Individual weights are calculated using a prior.
4. The remaining mass is split to constant and neighbor hypotheses, and the descriptor is normalized so all weights sum to 1.

Meaningful results are also obtained when the grid does not align perfectly with the colored rectangles:



experiments

experiment: level of discontinuity

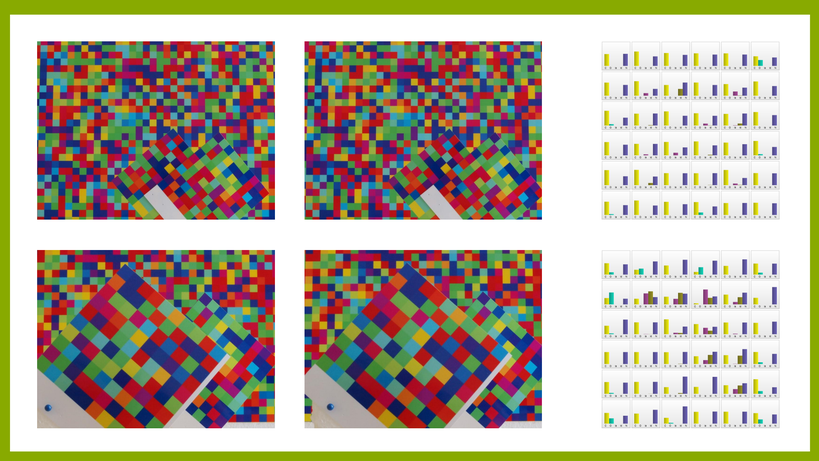
Goal: Investigate whether COIN descriptors can be used in assesing the amount of discontinuity around an interest point.
Setup: We use the sequence #014 from the DTU Robot Data Set. A number of interesting points are selected and their COIN descriptors calculated. For each COIN, we sum all inner changes of all patches at time t_2, calculating a discontinuity measure:

$$\Sigma_I^{t_2} = \sum \frac{I_a + I_d}{2}$$

	small discontinuity			large discontinuity			virtual crossing		coplanarity	
ID	3	4	7	1	8	9	6	5	2	10
$\bar{\Sigma}_I$	0.20	0.19	0.23	0.62	0.63	0.52	0.71	0.39	0.00	0.04

experiment: real vs. virtual crossings

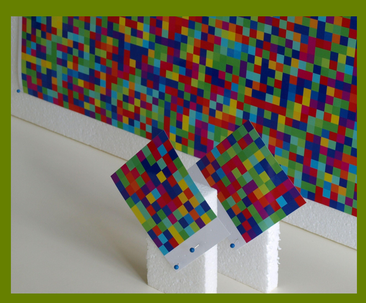
Goal: Investigate whether COINs of real crossings differ from COINs of real crossings.
Setup: Two sequences were taken, one with a real overlap between the two targets and another with a virtual overlap. COIN computation is dependant on the temporal direction, forward corresponding with "inner appearing", and backward corresponding with "inner disappearing". We visualize both "inner appearing" (I_a) and "inner disappearing" (I_d).



color world

Motivation: move away from synthetic examples to real images, but still keep a very controlled setting.

Goal: explore interesting scenarios, such as virtual and real crossings of targets, or holes in targets.



We acknowledge the support of Research Centre for Advanced Cooperative Systems (EU FP7 \#285939) and the University of Zagreb Development Fund. This poster was prepared for the ARTEMIS workshop of the 12th European Conference on Computer Vision (ECCV), October 7 - 13 2012, Florence, Italy.