APPEARANCE BASED OBJECT TRACKING IN STEREO SEQUENCES

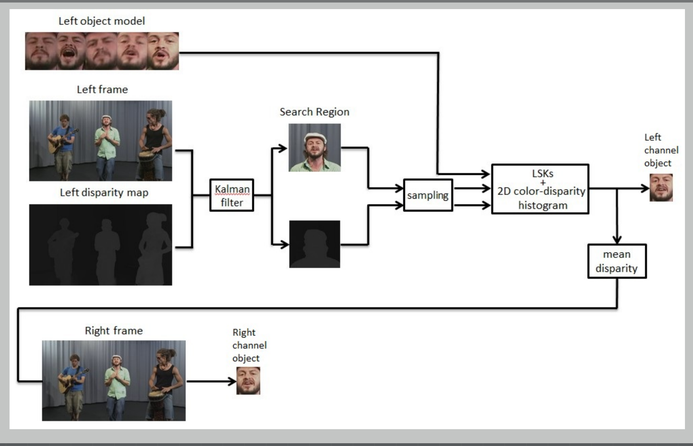
Olga Zoidi, Nikos Nikolaidis and Ioannis Pitas

Department of Informatics, Aristotle University of Thessaloniki, GREECE
Email: {ozoidi,nikolaid,pitas}@aiia.csd.auth.gr

1. Introduction

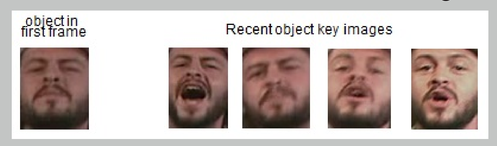
- A novel method for visual object tracking in stereo videos is proposed
- It requires no information about the camera calibration parameters
- It exploits low quality disparity maps extracted by a real-time disparity estimation algorithm
- Two representation methods for describing the object texture.
  - Color - disparity histograms
  - Local Steering Kernel (LSK) descriptors

2. Visual Object Tracking Overview



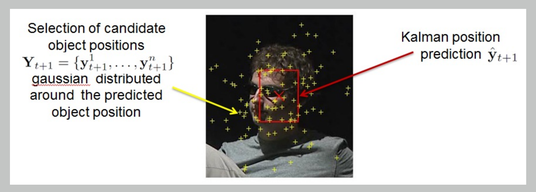
3. Object model

- Two object models are constructed for the left and right channel



4. Candidate object ROIs extraction

- The object position at frame $t+1$ is predicted with a 1^{st} order Kalman filter



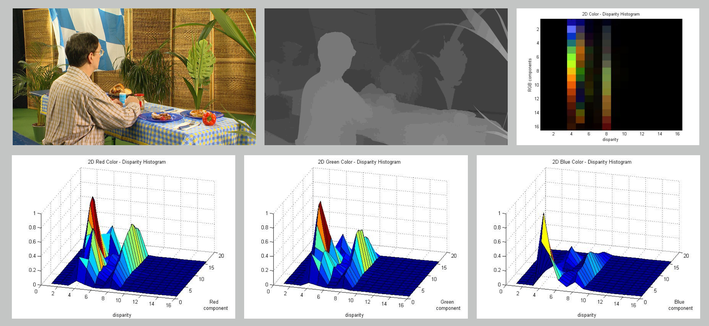
- Search region subsampling by selecting randomly n candidate object ROIs according to

$$\mathbf{Y}_{t+1} = \{\mathbf{y}^1_{t+1}, \ldots, \mathbf{y}^n_{t+1}\} \sim N(\hat{\mathbf{y}}_{t+1}, \Sigma),$$

where $\Sigma = \text{diag}[S_x/m, S_y/m]$, $S_x \times S_y$ the search region dimensions, $m = 4$

5. Color-disparity histograms

- 2D color-disparity histograms are constructed



- 80% of the candidate object ROIs with the lowest 2D-CDH similarity to the object at frame t (cosine similarity) are discarded
- The remaining 20% are compared to the object model with LSK similarity

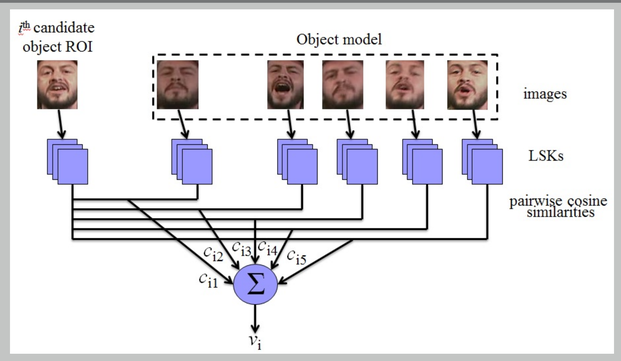
6. Local Steering Kernel feature extraction

- Local Steering Kernel (LSK) descriptors determine the similarity of an image pixel with its surrounding $P \times P$ pixels
- LSKs are computed for each pixel $\mathbf{p}$ by:

$$K(\mathbf{p}_l - \mathbf{p}) = \frac{\sqrt{\det(\mathbf{C}_l)}}{2\pi} \cdot \exp\left\{-\frac{(\mathbf{p}_l - \mathbf{p})^T \mathbf{C}_l (\mathbf{p}_l - \mathbf{p})}{2}\right\}, \quad l = 1, \ldots, P^2,$$

- where $\mathbf{C}_l$ is the covariance matrix of the gradient vectors of the image in a $P \times P$ window around $\mathbf{p}_l$
- LSKs become invariant to brightness and contrast changes by L_1 normalization
- Perform PCA to keep the d principal components

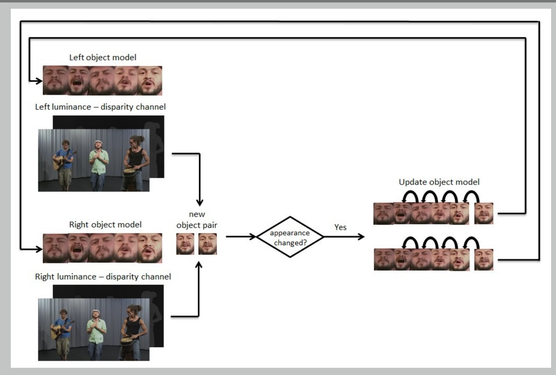
7. Object position detection at frame $t+1$



$$v_i = \lambda \frac{c_{i1}^2}{1 - c_{i1}^2} + \frac{1-\lambda}{k} \sum_{j=2}^{k} \frac{c_{ij}^2}{1 - c_{ij}^2} \in [0, +\infty)$$

- increased weight is given to the object ROI in the first frame
- the object ROI with the maximum v_i is the new object position

8. Object model update



9. Tracking Results

stereo tracker | Single-channel trackers

	length	stereo LSK tracker	monocular LSK tracker	CH tracker	L1 tracker	CT tracker	MIL tracker
video 1	930	**0.6324**	0.6069	0.2882	0.5580	0.3646	0.5284
video 2	629	**0.5633**	0.5313	0.0555	0.0994	0.4320	0.1077
video 3	689	**0.7136**	0.4671	0.4877	0.3422	0.3064	0.5776
video 4	500	0.6737	0.6962	0.6120	0.5901	0.5610	**0.7537**
video 5	500	0.6574	0.5498	0.4754	**0.6975**	0.5481	0.3266
video 6	500	**0.6808**	0.5236	0.4940	0.3653	0.1380	0.4322
video 7	165	**0.7554**	**0.7558**	0.6440	0.1568	0.5386	0.3881
video 8	95	**0.5187**	0.2881	0.1938	0.0542	0.1835	0.1983
video 9	545	**0.5993**	0.4972	0.3318	0.0242	0.5528	0.5611
OTA		**0.6459**	0.5553	0.3811	0.3707	0.4070	0.4617
ATA variance		**0.0029**	0.0074	0.0299	0.0526	0.0189	0.0345

OTA: Maximum overlap between ground truth and tracking result

10. Conclusions

- The proposed stereo tracker is successful in tracking rigid objects under pose changes, small rotation changes and small scale changes
- It outperforms state of the art monocular appearance based trackers

11. Acknowledgement

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 287674 (3DTVS)