Contact person: Christel.Chamaret@technicolor.com

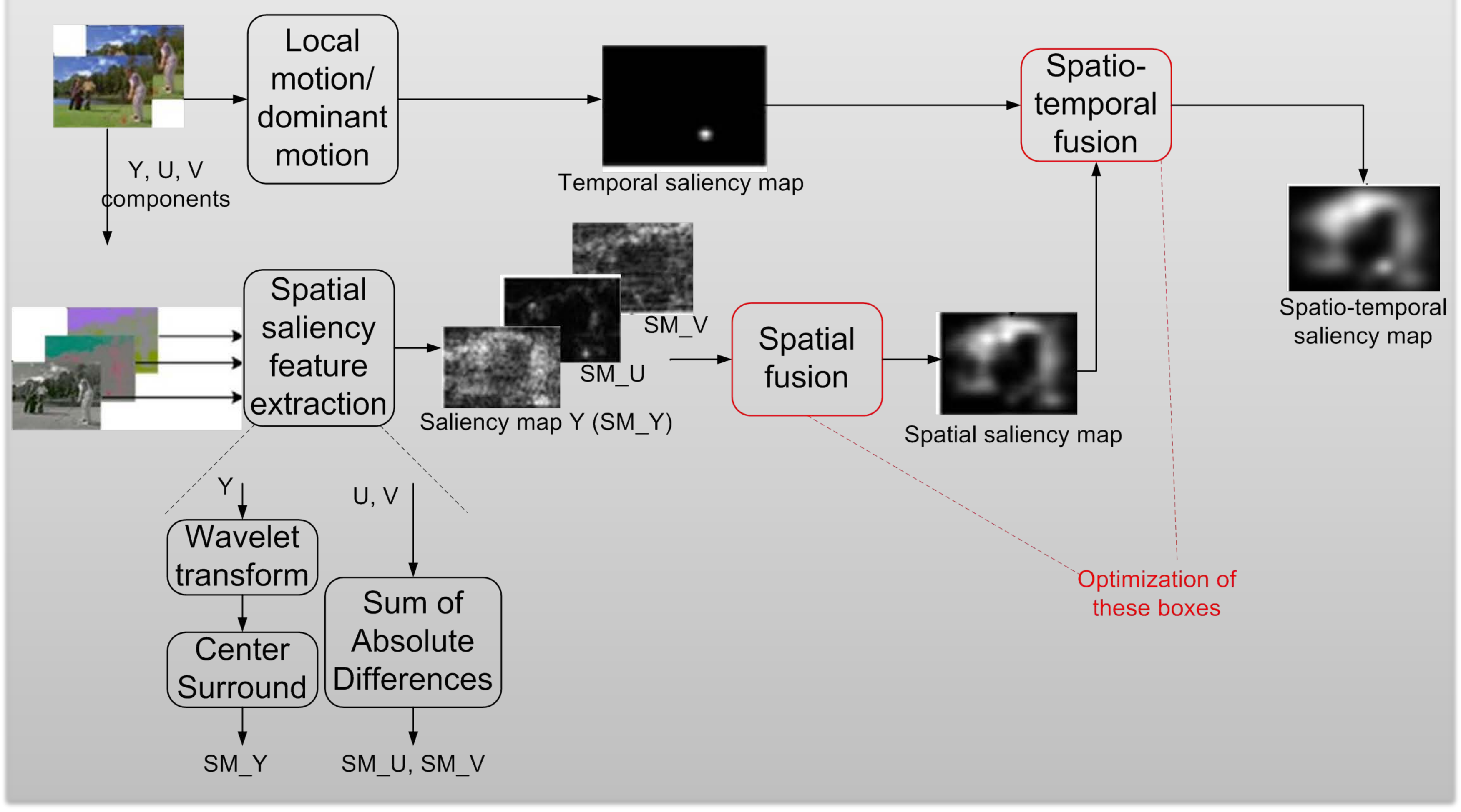
Spatio-temporal combination of saliency maps and eye-tracking assessment of different strategies

Christel Chamaret, Olivier Le Meur and Jean-Claude Chevet

Issue: assessment of sophisticated fusion schemes of saliency maps

Context

Visual attention model, fusion of uncorrelated data



Validation methodology

Eye-tracking experiments: user fixation points

Still images

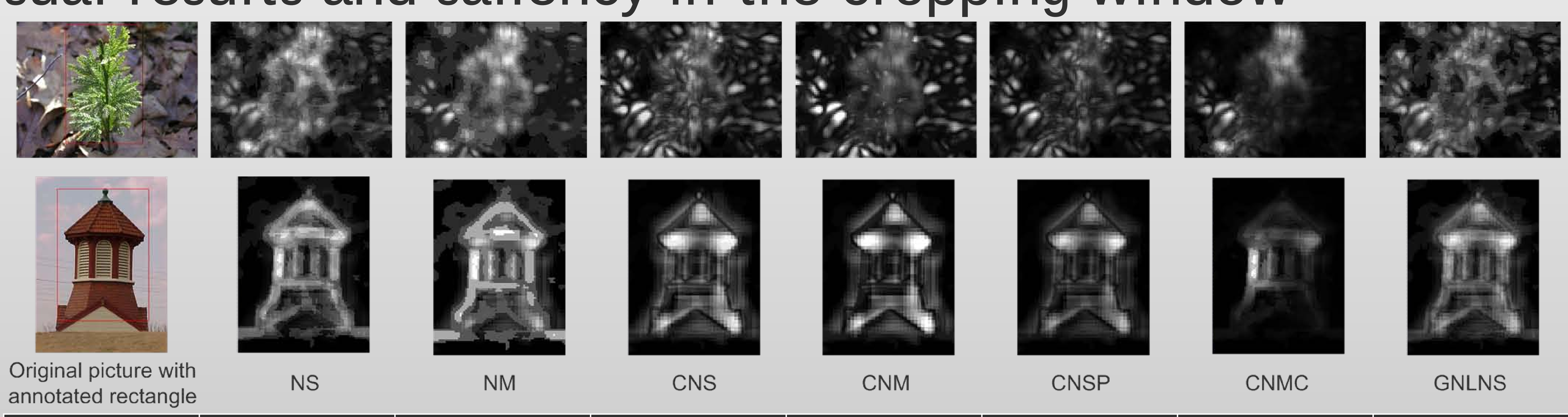
Video sequences

Saliency quantity in annotated rectangle

CC, KL, ROC metrics

Still images

Visual results and saliency in the cropping window



	NS	NM	CNS	CNM	CNSP	CNMC	GNLNS
Mean	52.38	50.28	52.52	51.18	53.36	71.05	52.79
St.Dev	19.42	18.78	19.07	19	19.24	25.07	19.29
Min.	7.29	8.34	8.64	9.06	8.5	0	7.49
Max.	98.95	98.36	98.26	97.61	98.39	100	98.62

Statistical results with video sequences

Operators	NS			NM			CNS			CNM			CNSP			CNMC			GNLNS		
Metrics	CC	KL	ROC	CC	KL	ROC	CC	KL	ROC	CC	KL	ROC	CC	KL	ROC	CC	KL	ROC	CC	KL	ROC
Canoa	0.44	11.5	0.71	0.44	12.1	0.70	0.44	11.5	0.71	0.44	12	0.70	0.55	10.3	0.84	**0.61**	**8.5**	**0.87**	0.59	8.5	0.84
Kayak	0.48	12.3	0.76	0.44	12.9	0.73	0.49	12.2	0.76	0.44	12.9	0.73	0.59	10.8	0.86	**0.63**	**8.6**	**0.87**	0.57	9.6	0.82
PatVit	**0.65**	**7.6**	**0.85**	0.61	9.3	0.83	0.65	7.8	0.85	0.57	10	0.81	0.61	8.6	0.83	0.6	8.5	0.83	0.52	10.4	0.77
Stefan	0.53	9.9	0.8	0.48	11.1	0.79	0.52	9.8	0.8	0.5	10.6	0.8	0.53	9.7	0.81	**0.54**	**9**	**0.82**	0.49	10.1	0.8
Table	0.54	9.3	0.8	0.52	9.9	0.79	0.54	9.3	0.8	0.52	9.9	0.79	0.55	9	0.8	**0.57**	**8.4**	**0.8**	0.53	9.9	0.8
Titleist	**0.62**	8.5	0.85	0.6	9.4	0.84	0.62	8.8	**0.85**	0.57	10.1	0.83	0.6	9	0.84	0.61	**8.1**	0.85	0.55	10.2	0.81
Mean	0.55	9.7	0.8	0.52	10.6	0.78	0.55	9.7	0.8	0.51	10.8	0.78	0.58	9.4	0.83	**0.6**	**8.2**	**0.85**	0.54	9.6	0.81
St. Dev.	0.07	1.5	0.04	0.06	1.3	0.04	0.06	1.4	0.04	0.04	1.1	0.04	0.03	0.8	0.02	0.03	0.7	0.02	0.02	0.7	0.02

Fusion operators: from the simplest to the most complex one

Operator	Formula	Description	Advantage	Weakness
NS: Normalized and Sum	$Out = N\left(\sum_i N(SM_i)\right)$	Normalization of maps to the same dynamic range [0-1] .	Computational simplicity.	Promotion of irrelevant information due to a noisy map. Loss of the relative importance between the maps. No spatial competition.
NM: Normalized and Maximum	$Out = \max_i\left(N(SM_i)\right)$	Maximum operator instead of summation (compared to NS).		Same as NS
CNS: Coherent Normalization and Sum	$Out = \sum_i N_c(SM_i)$	Empirical determination of the maximum saliency value for each visual dimension. Use of particular pictures: high contrasted color, black pictures…	Conservation of the relative importance between maps.	Not biologically plausible because this method tends to favor a unique location per map.
CNM: Coherent Normalization and Maximum	$Out = \max_i\left(N_c(SM_i)\right)$	Maximum operator instead of summation (compared to CNS).		Same as CNS
CNSP: Coherent Normalization, Sum plus Product	$Out = \sum_i N_c(SM_i) + \prod_i \left(1 + N_c(SM_i)\right)$	Highlight of local inter-map redundancies (via product operator) and incoherences (via summation operator).	Competition between the different maps. Promotion of an item that generates saliency in several dimensions.	No intra-map competition.
CNMC: Coherent Normalization, intra and inter Map Competition	$Out = \sum_i \widehat{SM}_i + \prod_i \left(1 + \widehat{SM}_i\right)$ $\widehat{SM}_i = \frac{N_c(SM_i)}{NEAREST\left(WTA\left(N_c(SM_i)\right)\right)}$	Upgrade of CNSP approach by using a WTA algorithm with localized inhibitory spread. Detection of the local maximum and use of them to locally favor some parts of the picture. The number of maximum peaks, their values and the difference value between two consecutive maximum peaks are required to keep only the most interesting areas.	Interesting approach because a sparse distribution of maximum is computed in order to favor and to promote certain locations in the scene. This property is close to the biological behavior.	Several thresholds to drive the algorithm: the first one for avoiding the weak saliency peaks, a second one is the ratio between two consecutive maximum determining the numbers of local maximum and the third one test the relevance of the second threshold.
GNLNS: Global Non-Linear Normalization followed by Normalization	$Out = \sum_i \left[\left(N(SM_i)\right) \cdot \left(M_i - m_i\right)^2\right]$	Implemented by L. Itti. Promotion of the maps having few saliency peaks and decrease of the impact of the maps having an uniform distribution and a lot of saliency peaks.		Highly sensitive to noise in the maps due to the global normalization.

Conclusion: best results when including intra and inter-map competition

technicolor