Spatio-Temporal Quasi-Flat Zones for Morphological Video Segmentation

Jonathan Weber † Sébastien Lefèvre ‡ Pierre Gançarski †

† LSIIT, University of Strasbourg, France
‡ VALORIA, University of South Brittany, France

Context

Video segmentation issues:

- Large amount of data to be segmented (in the context of video mining)
- Low computation time is required (for interactive segmentation)
- Segmentation is user-dependent: personalization is needed

Proposed solutions:

- Efficient segmentation relying on an offline reduction of the data space
- User-driven segmentation with iterative feedback

Quasi-Flat Zones

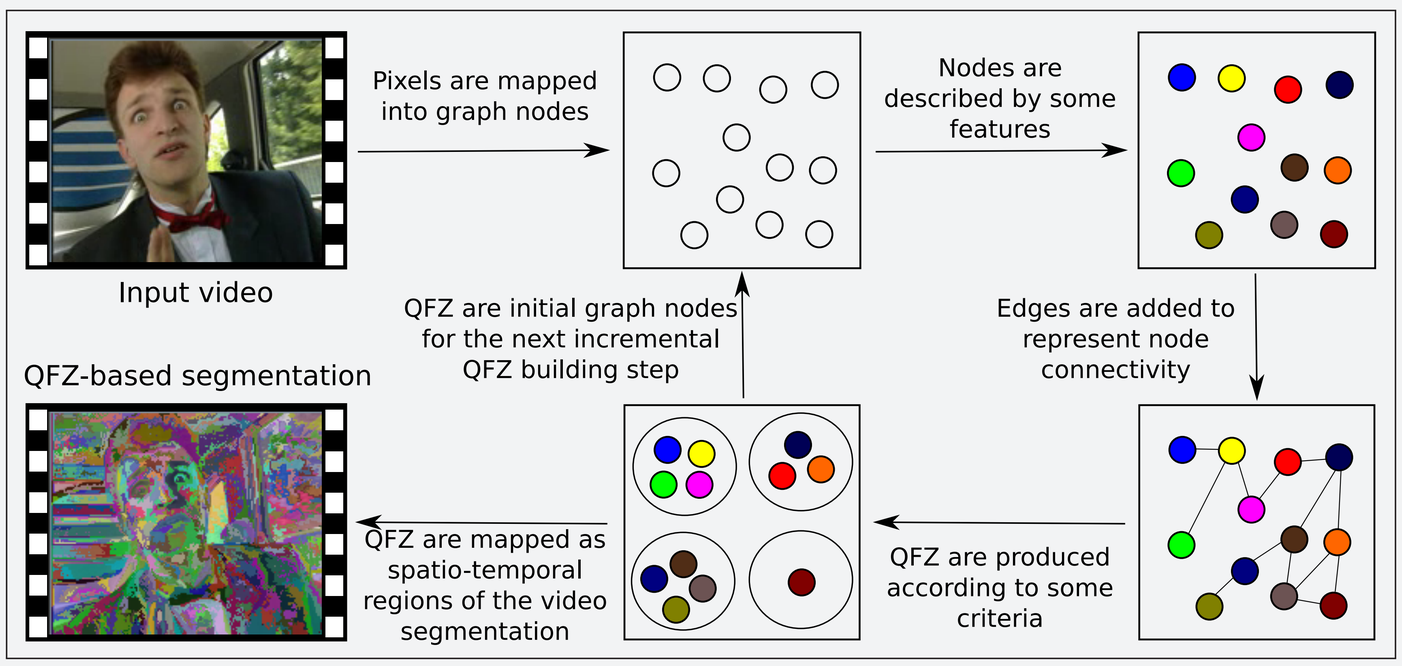
Quasi-Flat Zones (QFZ):

- Connected areas of homogeneous pixels
- 2 homogeneity criteria are mainly used:
  - Local range (α)
  - Global range (ω)

QFZ definitions were unified by Soille [PAMI 2008] under constrained connectivity (using logical predicates).

1.Extension of QFZ to video data

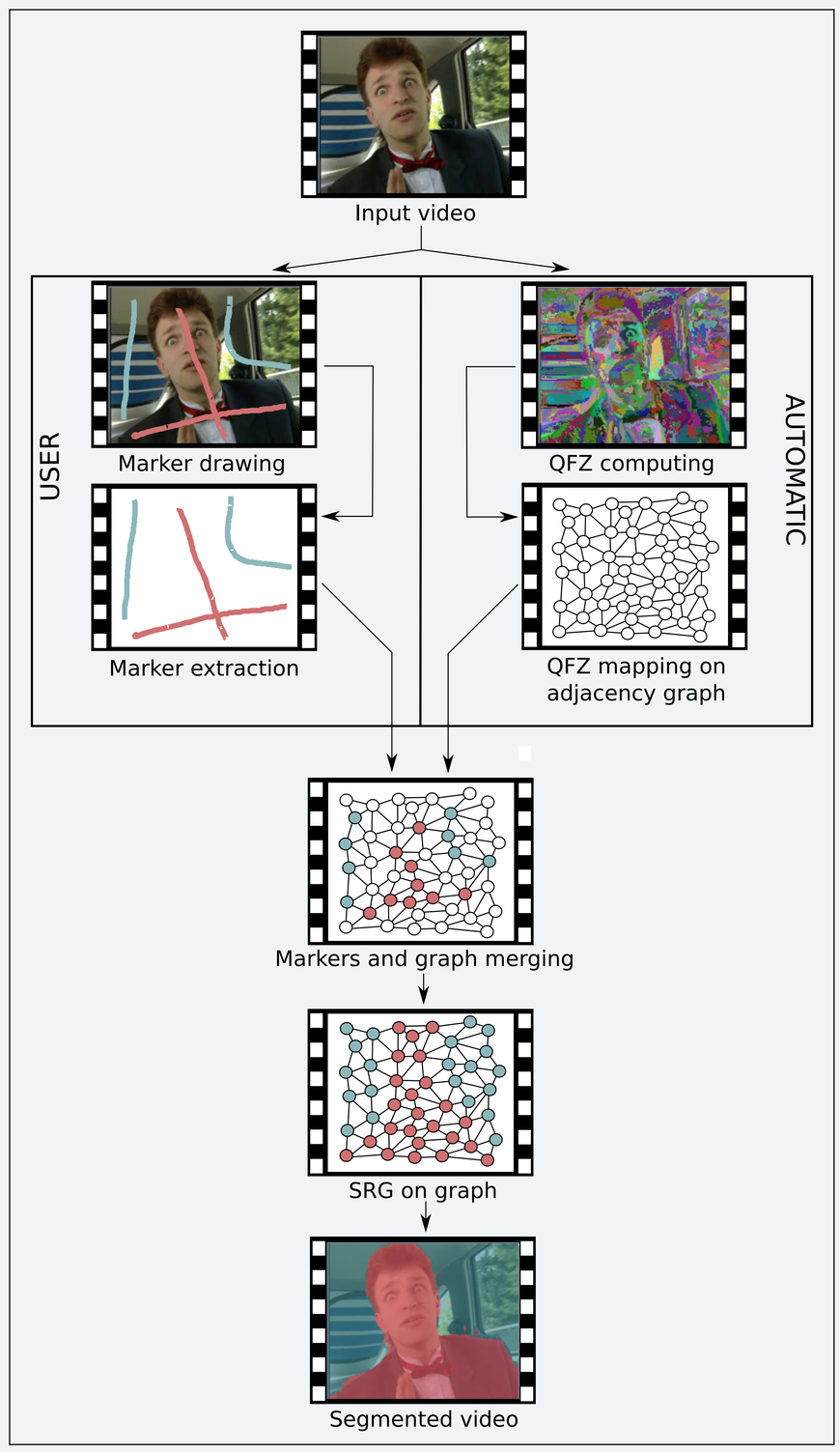
- Straight extension (video = 3D volume) is not adapted
- Sequential processing of spatial and temporal dimensions is more relevant
  $\Rightarrow$ induces an incremental approach for QFZ production



- Filtering criterion to reduce amount of QFZs: spatial mean area ($area^*$)
- Result is an efficient reduction of the data space, e.g., *carphone* extract
  - 9 656 064 pixels in the input video
  - parameters: $\alpha = \omega = 20$, $area^* = 10$
  - 30 671 QFZ with $2D + t$ approach
  - 4 713 QFZ with $t + 2D$ approach

2.Marker-Based Quasi-Flat Zones

- 2-step personalized segmentation:
  (a) Production of spatio-temporal QFZ (offline)
  (b) Objects of interest are marked by the user and QFZ are subsequently merged according to these markers (online)



3.Experiments and results

Method	α, ω	$area^*$	Mean Jaccard-Index			
			carphone		*foreman*	
			Set 1	Set 2	Set 1	Set 2
MBQFZ 2D+t	30	10	0.782	0.905	0.710	**0.952**
	50	50	**0.825**	0.910	0.674	0.884
	90	50	0.793	0.908	**0.791**	0.859
MBQFZ t+2D	20	60	0.767	**0.928**	0.695	0.944
	40	100	0.749	0.925	0.656	0.940
	100	70	0.781	0.919	0.637	0.935
SRG	-	-	0.641	0.548	0.529	0.400
MBWS	-	-	0.749	0.897	0.634	0.946

Results obtained on *carphone* extract (80 frames) and *foreman* extract (40 frames) with two sets of markers only defined on the median frame.

Method	α, ω	# QFZ	Computing time in **ms**		
			Offline	Online (per frame)	
MBQFZ 2D+t	10	28 612	44 390	528	(1.39)
	20	30 671	35 510	550	(1.44)
	30	27 713	38 762	508	(1.33)
$area^* = 10$	40	22 202	43 280	364	(0.96)
	50	18 501	46 343	326	(0.86)
MBQFZ t+2D	10	3 772	44 781	108	(0.28)
	20	4 713	32 080	123	(0.32)
	30	4 649	26 957	116	(0.30)
$area^* = 10$	40	3 842	26 128	107	(0.28)
	50	3 147	25 133	98	(0.26)
SRG	–	–	0	56 636	(148.65)
MBWS	–	–	3 354	17 312	(45.44)

Offline and online computation times required to process the whole *carphone* sequence (9 656 064 pixels, 176 x 144 on 381 frames).

Conclusion and Perspectives

Interactive video segmentation driven by QFZ:

- efficient reduction of the input data space
- personalized segmentation from iterative user feedback
- low online computational cost during iterative steps

Future works:

- improve the user feedback (marker refinement)
- application to other data spaces (e.g., optical flow)
- extension to cosegmentation (i.e., segment a complete video dataset from a few user-driven segmentations)

ISMM 2011
10th Int. Symposium on Mathematical Morphology
6th-8th July 2011 - Intra, Lake Maggiore, Italy