Interactive Video Segmentation based on Quasi-Flat Zones

Jonathan Weber, Sébastien Lefèvre, Pierre Gançarski

ISPA 2011
7th International Symposium on Image and Signal Processing and Analysis
Dubrovnik, Croatia

September 5th, 2011

Context: Explosion of available video data volume

Mass of video data

- Nowadays
  - YouTube serves up 3 billions video clips per day
- In the future (according to CISCO)
  - 2012: video will represent 50% of Internet traffic
  - 2015: 100 millions minutes of video will be broadcasted each second

Dealing with these data

- Often need prior segmentation
  - To use elements more semantically meaningful than pixels (regions, objects ...)

Context: Explosion of available video data volume

Mass of video data

- Nowadays
  - YouTube serves up 3 billions video clips per day
- In the future (according to CISCO)
  - 2012: video will represent 50% of Internet traffic
  - 2015: 100 millions minutes of video will be broadcasted each second

Dealing with these data

- Often need prior segmentation
  - To use elements more semantically meaningful than pixels (regions, objects ...)

Context: Explosion of available video data volume

Mass of video data

- Nowadays
  - YouTube serves up 3 billions video clips per day
- In the future (according to CISCO)
  - 2012: video will represent 50% of Internet traffic
  - 2015: 100 millions minutes of video will be broadcasted each second

Dealing with these data

- Often need prior segmentation
  - To use elements more semantically meaningful than pixels (regions, objects ...)

Context: Explosion of available video data volume

Mass of video data

- Nowadays
  - YouTube serves up 3 billions video clips per day
- In the future (according to CISCO)
  - 2012: video will represent 50% of Internet traffic
  - 2015: 100 millions minutes of video will be broadcasted each second

Dealing with these data

- Often need prior segmentation
  - To use elements more semantically meaningful than pixels (regions, objects ...)

Context: Problems induced by video segmentation

Video represents a huge data volume

- 1 HD video second $\simeq$ 178 MB (uncompressed)
  - $\Rightarrow$ High computationnal cost
  - $\Rightarrow$ High memory cost

Semantic gap

- Difference between numeric computer representation and human-being video content interpretation
  - $\Rightarrow$ Automatic video segmentation is a complex task

Context: Problems induced by video segmentation

Video represents a huge data volume

- 1 HD video second $\simeq$ 178 MB (uncompressed)
  - $\Rightarrow$ High computationnal cost
  - $\Rightarrow$ High memory cost

Semantic gap

- Difference between numeric computer representation and human-being video content interpretation
  - $\Rightarrow$ Automatic video segmentation is a complex task

Context: Problems induced by video segmentation

Video represents a huge data volume

- 1 HD video second $\simeq$ 178 MB (uncompressed)
  - $\Rightarrow$ High computationnal cost
  - $\Rightarrow$ High memory cost

Semantic gap

- Difference between numeric computer representation and human-being video content interpretation
  - $\Rightarrow$ Automatic video segmentation is a complex task

Context: Proposed solutions

Data volume reduction

- Apply video segmentation process on an oversegmentation of the video using quasi-flat zones and not anymore on pixels

User involvment

- Interactively involve user in the segmentation process to bridge the semantic gap

Context: Proposed solutions

Data volume reduction

- Apply video segmentation process on an oversegmentation of the video using quasi-flat zones and not anymore on pixels

User involvment

- Interactively involve user in the segmentation process to bridge the semantic gap

Context: Proposed solutions

Data volume reduction

- Apply video segmentation process on an oversegmentation of the video using quasi-flat zones and not anymore on pixels

User involvment

- Interactively involve user in the segmentation process to bridge the semantic gap

Quasi-Flat Zones for still images

Flat zones (connected components)

- Flat zones: connected set of pixels with same values
- Extreme oversegmentation

Quasi-Flat zones: Generalization of flat zones

- Quasi-Flat zones: connected set of pixels with close values
- Closeness depends on parameters
  - local range α: maximum difference between neighbouring pixels values
  - global range ω: maximum difference between all pixels values
- Keep the relative good frontiers from flat zones

Quasi-Flat Zones for still images

Flat zones (connected components)

- Flat zones: connected set of pixels with same values
- Extreme oversegmentation

Quasi-Flat zones: Generalization of flat zones

- Quasi-Flat zones: connected set of pixels with close values
- Closeness depends on parameters
  - local range α: maximum difference between neighbouring pixels values
  - global range ω: maximum difference between all pixels values
- Keep the relative good frontiers from flat zones

Quasi-Flat Zones for still images

Flat zones (connected components)

- Flat zones: connected set of pixels with same values
- Extreme oversegmentation

Quasi-Flat zones: Generalization of flat zones

- Quasi-Flat zones: connected set of pixels with close values
- Closeness depends on parameters
  - local range α: maximum difference between neighbouring pixels values
  - global range ω: maximum difference between all pixels values
- Keep the relative good frontiers from flat zones

Quasi-Flat Zones for still images

Flat zones (connected components)

- Flat zones: connected set of pixels with same values
- Extreme oversegmentation

Quasi-Flat zones: Generalization of flat zones

- Quasi-Flat zones: connected set of pixels with close values
- Closeness depends on parameters
  - local range α: maximum difference between neighbouring pixels values
  - global range ω: maximum difference between all pixels values
- Keep the relative good frontiers from flat zones

Quasi-Flat Zones for still images

Flat zones (connected components)

- Flat zones: connected set of pixels with same values
- Extreme oversegmentation

Quasi-Flat zones: Generalization of flat zones

- Quasi-Flat zones: connected set of pixels with close values
- Closeness depends on parameters
  - local range α: maximum difference between neighbouring pixels values
  - global range ω: maximum difference between all pixels values
- Keep the relative good frontiers from flat zones

Quasi-Flat Zones for still images

Flat zones (connected components)

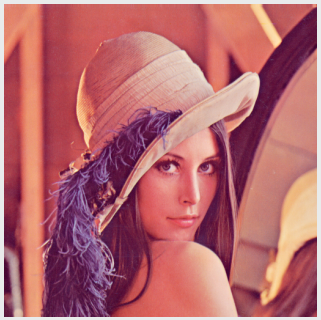
- Flat zones: connected set of pixels with same values
- Extreme oversegmentation

Quasi-Flat zones: Generalization of flat zones

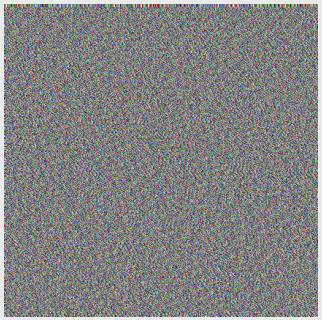
- Quasi-Flat zones: connected set of pixels with close values
- Closeness depends on parameters
  - local range α: maximum difference between neighbouring pixels values
  - global range ω: maximum difference between all pixels values
- Keep the relative good frontiers from flat zones

Quasi-Flat Zones for still images

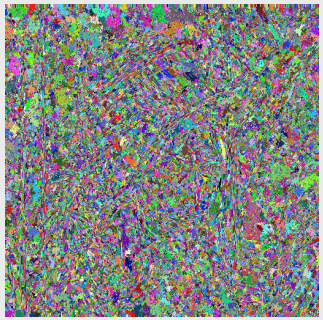
Results



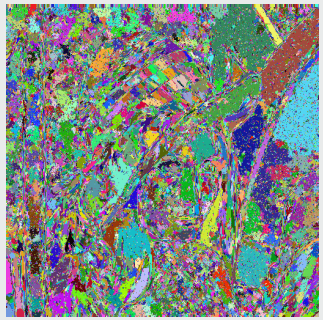
Original
262 144 pixels



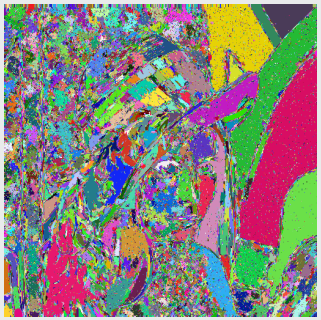
$\alpha = \omega = 0$
258 389 QFZ



$\alpha = \omega = 25$
82 239 QFZ



$\alpha = \omega = 50$
58 219 QFZ



$\alpha = \omega = 75$
45 695 QFZ

$\alpha = \omega = 100$
34 651 QFZ

$\alpha = \omega = 125$
29 444 QFZ

$\alpha = \omega = 150$
22 730 QFZ

Quasi-Flat Zones extended to Video

Straight 3D extension

- Considering a 3D neighbourhood (X,Y,T) instead of the 2D neighbourhood (X,Y) used for still images
  - \+ trivial extension
  - – weak results

2D+t/t+2D extension

- Produce QFZ sequentially on spatial (X,Y) and temporal (T) dimensions
  - \+ better results
  - – less trivial

Quasi-Flat Zones extended to Video

Straight 3D extension

- Considering a 3D neighbourhood (X,Y,T) instead of the 2D neighbourhood (X,Y) used for still images
  - \+ trivial extension
  - – weak results

2D+t/t+2D extension

- Produce QFZ sequentially on spatial (X,Y) and temporal (T) dimensions
  - \+ better results
  - – less trivial

Quasi-Flat Zones extended to Video

Straight 3D extension

- Considering a 3D neighbourhood (X,Y,T) instead of the 2D neighbourhood (X,Y) used for still images
  - \+ trivial extension
  - – weak results

2D+t/t+2D extension

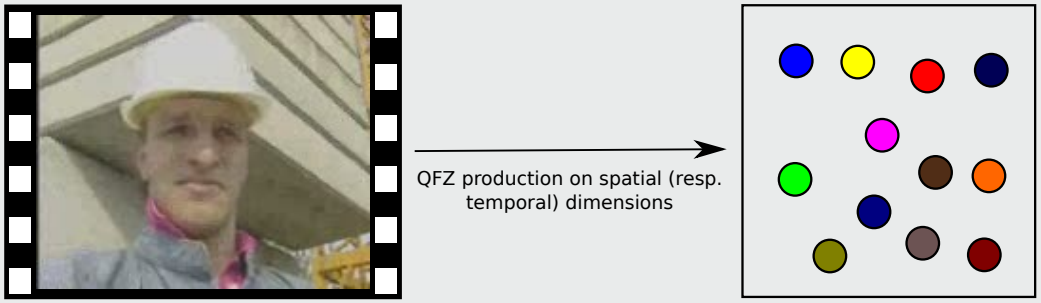
- Produce QFZ sequentially on spatial (X,Y) and temporal (T) dimensions
  - \+ better results
  - – less trivial

Quasi-Flat Zones extended to Video

2D+t/t+2D Quasi-flat zones framework

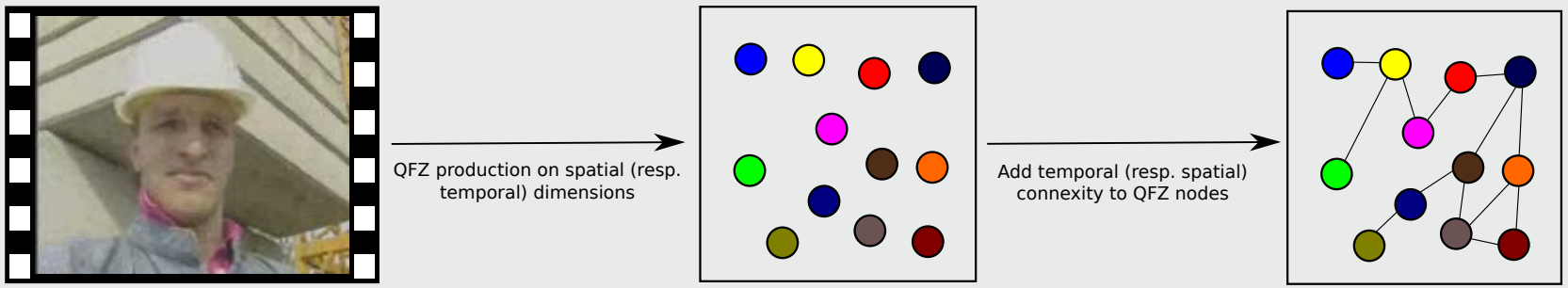
Quasi-Flat Zones extended to Video

2D+t/t+2D Quasi-flat zones framework



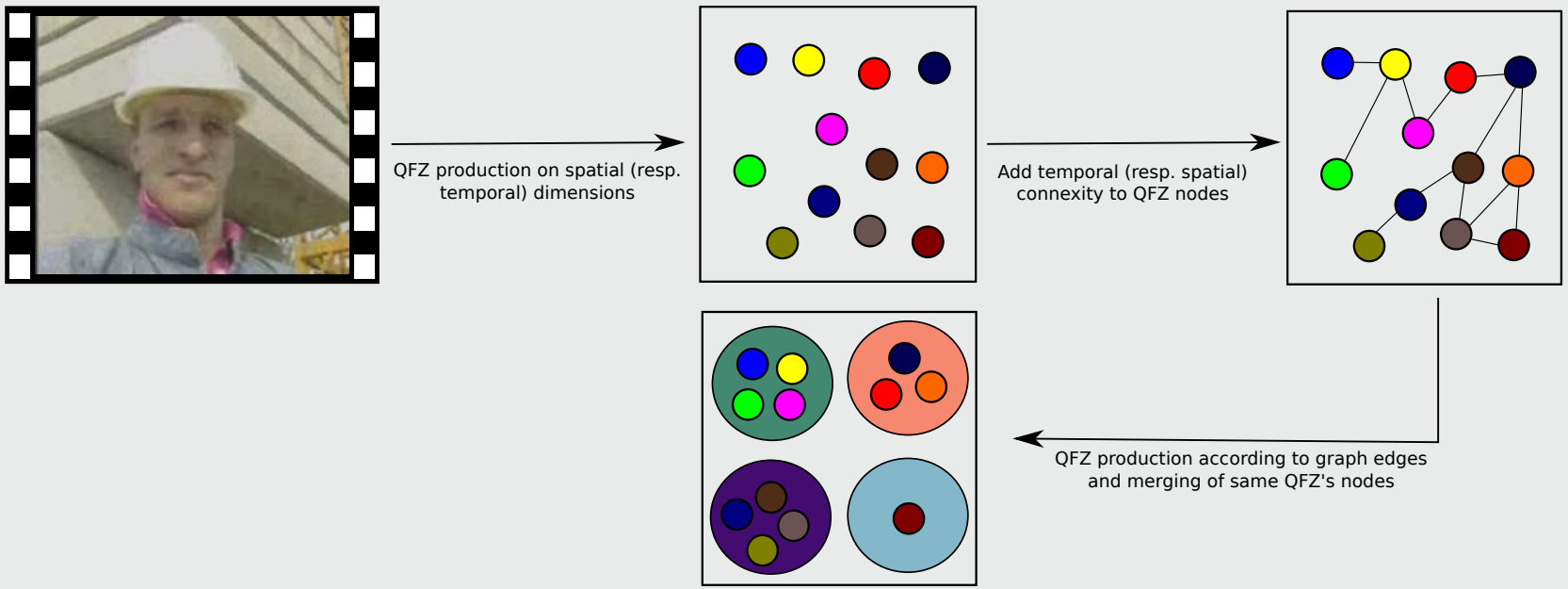
Quasi-Flat Zones extended to Video

2D+t/t+2D Quasi-flat zones framework



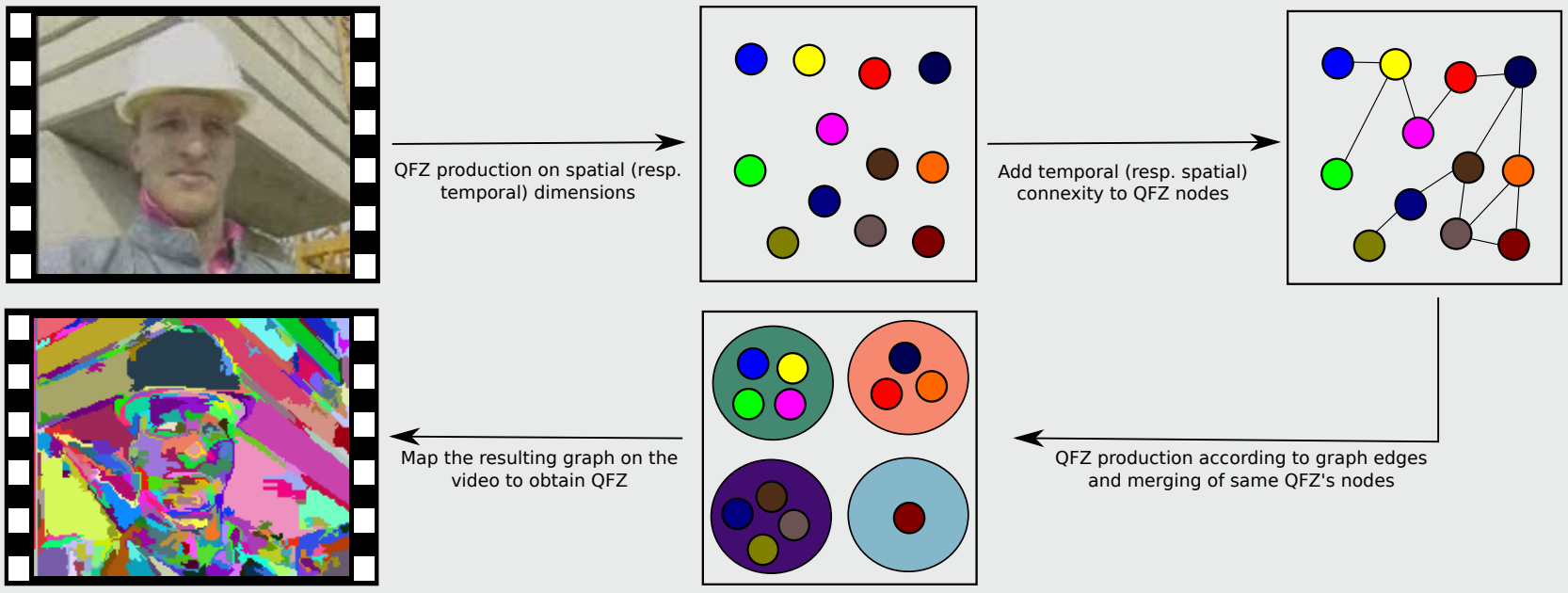
Quasi-Flat Zones extended to Video

2D+t/t+2D Quasi-flat zones framework



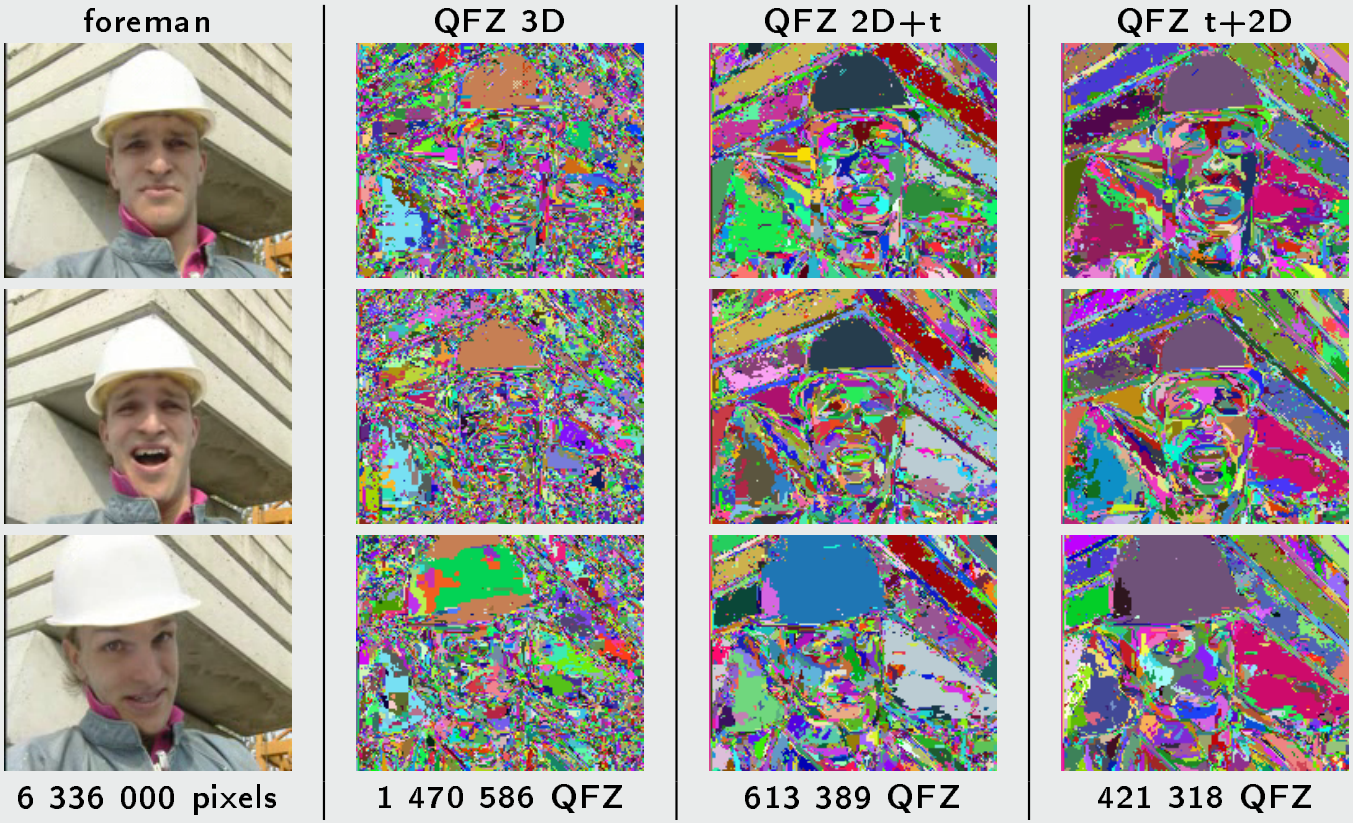
Quasi-Flat Zones extended to Video

2D+t/t+2D Quasi-flat zones framework



Quasi-Flat Zones extended to Video

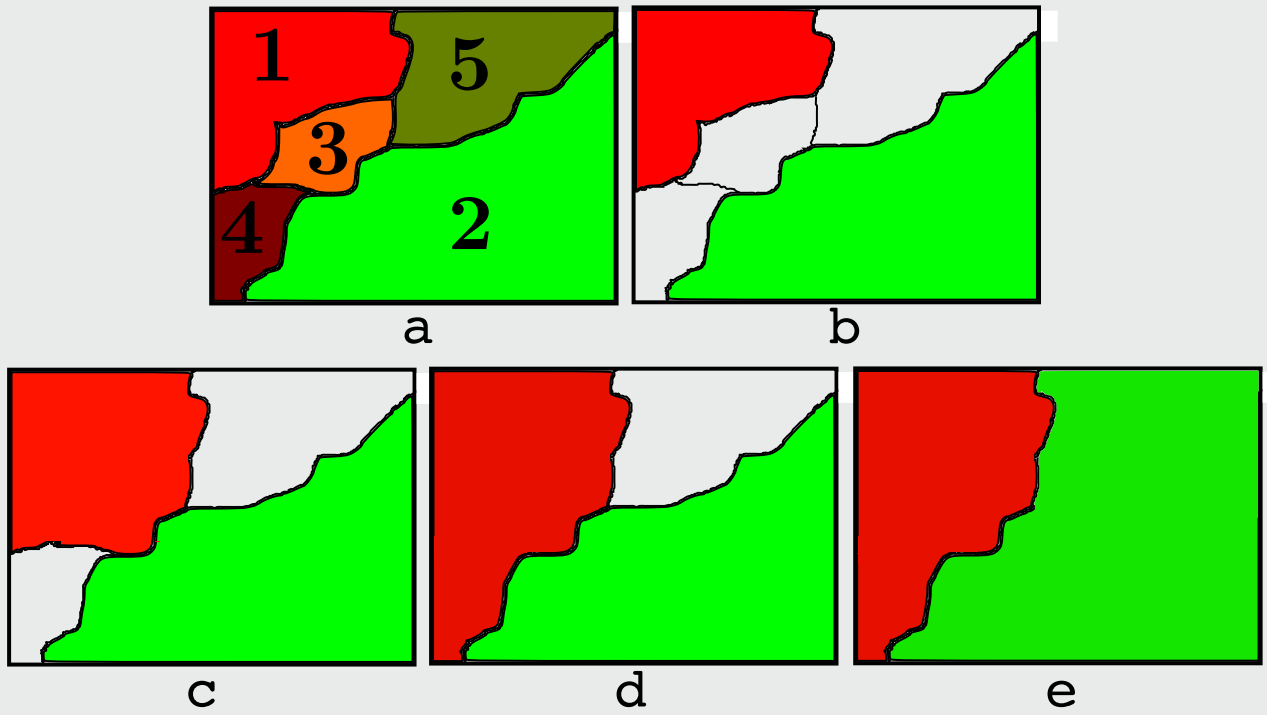
Results



Quasi-Flat Zones extended to Video

Oversegmentation reduction

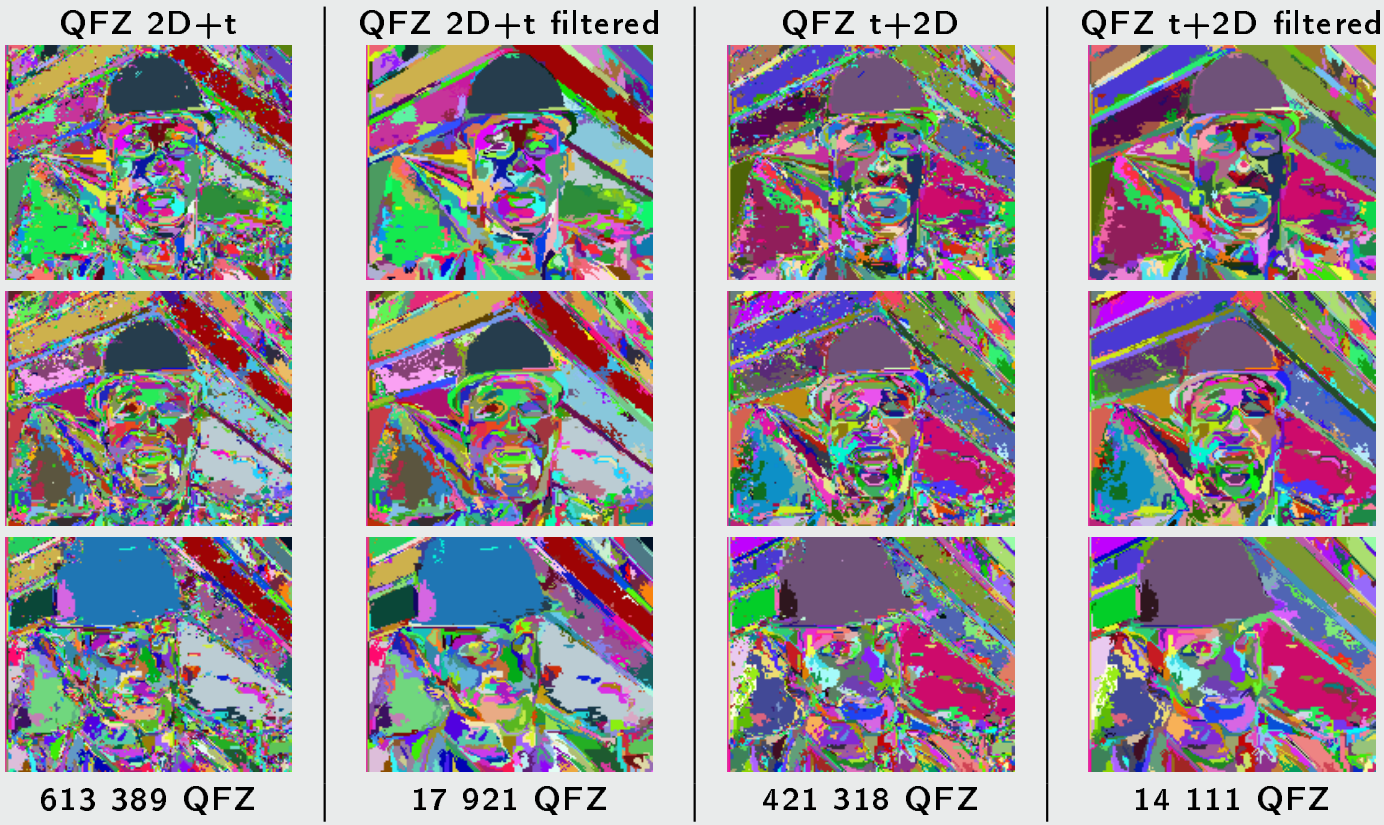
- Filtering QFZ
  - Criterion: Threshold on mean spatial size of QFZ
  - Reconstruction: Seeded Region Growing on QFZ



Quasi-Flat Zones extended to Video

Filtering results

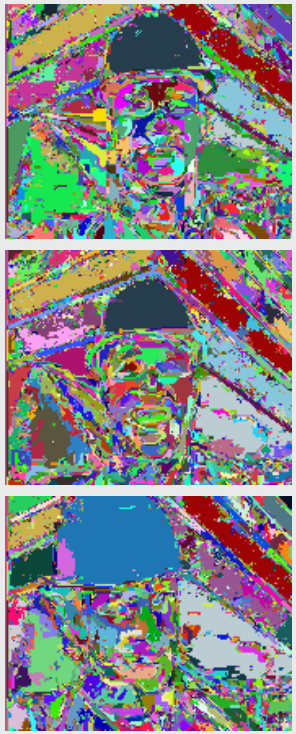
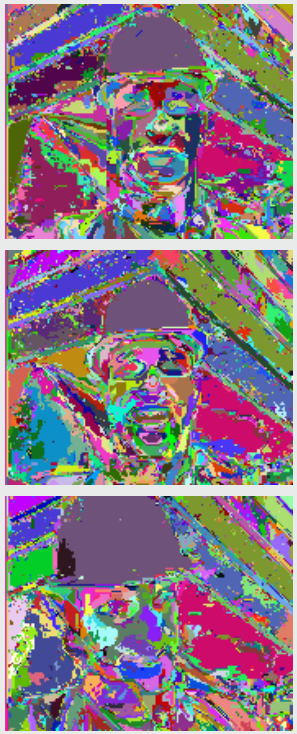
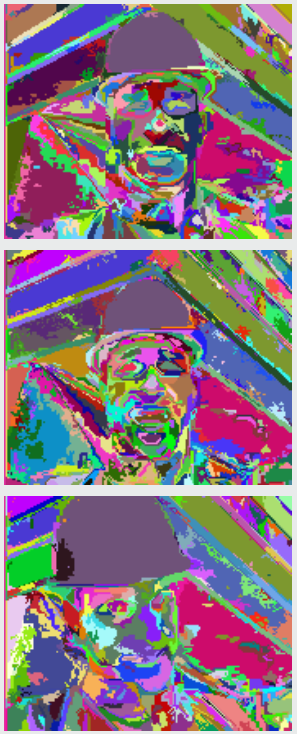
- Mean spatial area threshold: 10 pixels



Quasi-Flat Zones extended to Video

Filtering results

- Mean spatial area threshold: 10 pixels

QFZ 2D+t	QFZ 2D+t filtered	QFZ t+2D	QFZ t+2D filtered
			
613 389 QFZ	17 921 QFZ	421 318 QFZ	14 111 QFZ

Interactive Segmentation

Constraints

- Interaction has to be intuitive
- Requiring time for the user has to be minimal

Proposed solutions

- Marker-drawing interaction
- Interactive segmentation based on video QFZ
  - QFZ production made offline
  - Seeded Region growing algorithm used to merge QFZ according to markers

Interactive Segmentation

Constraints

- Interaction has to be intuitive
- Requiring time for the user has to be minimal

Proposed solutions

- Marker-drawing interaction
- Interactive segmentation based on video QFZ
  - QFZ production made offline
  - Seeded Region growing algorithm used to merge QFZ according to markers

Interactive Segmentation

Constraints

- Interaction has to be intuitive
- Requiring time for the user has to be minimal

Proposed solutions

- Marker-drawing interaction
- Interactive segmentation based on video QFZ
  - QFZ production made offline
  - Seeded Region growing algorithm used to merge QFZ according to markers

Interactive Segmentation

Constraints

- Interaction has to be intuitive
- Requiring time for the user has to be minimal

Proposed solutions

- Marker-drawing interaction
- Interactive segmentation based on video QFZ
  - QFZ production made offline
  - Seeded Region growing algorithm used to merge QFZ according to markers

Interactive Segmentation

Constraints

- Interaction has to be intuitive
- Requiring time for the user has to be minimal

Proposed solutions

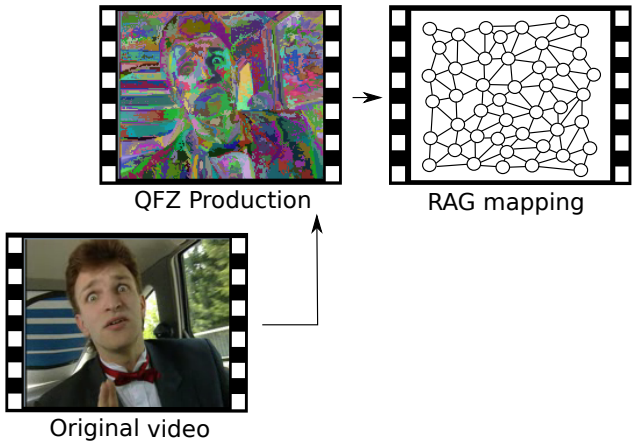
- Marker-drawing interaction
- Interactive segmentation based on video QFZ
  - QFZ production made offline
  - Seeded Region growing algorithm used to merge QFZ according to markers

Interactive Segmentation: Framework

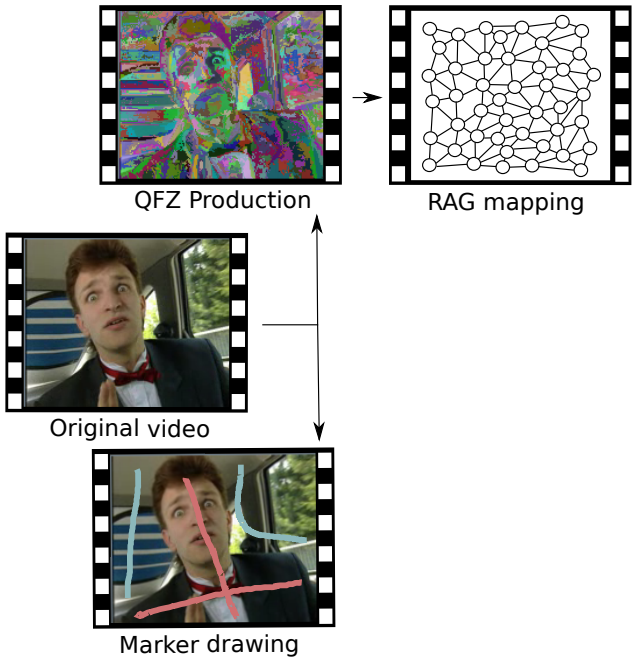
Original video

Interactive Segmentation: Framework

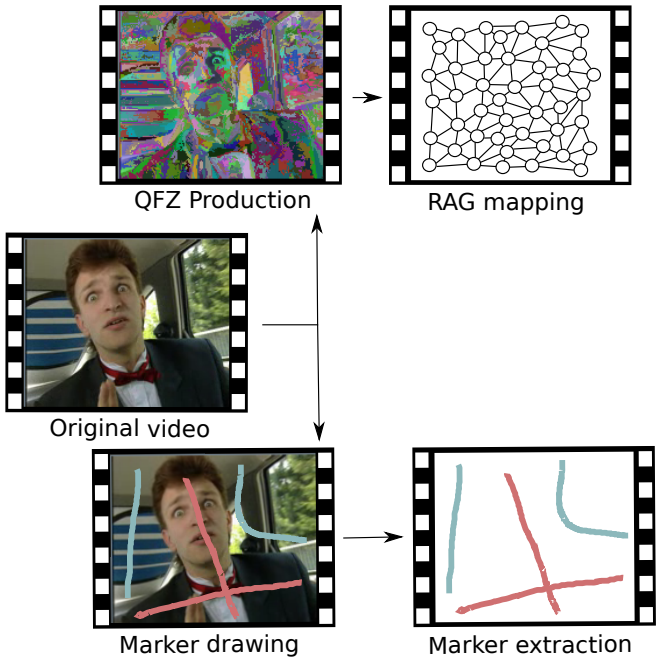
Interactive Segmentation: Framework



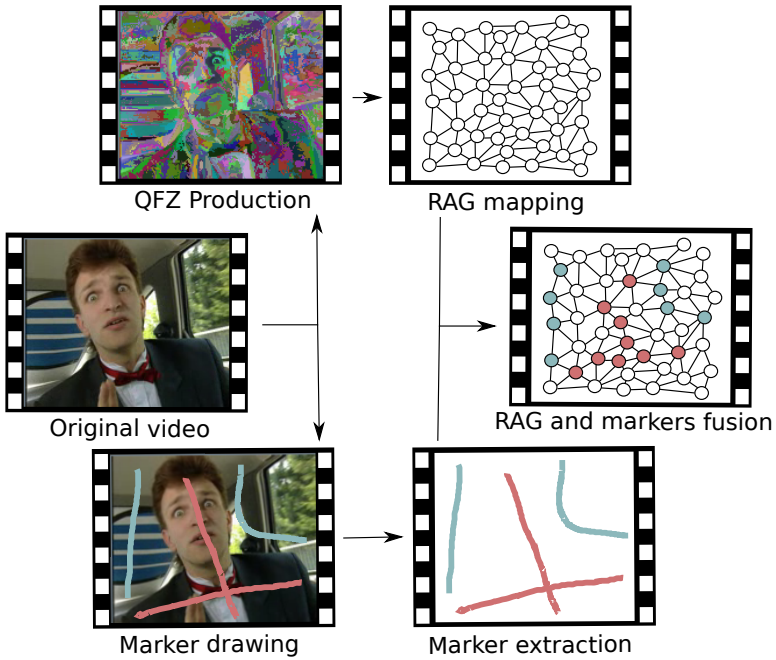
Interactive Segmentation: Framework



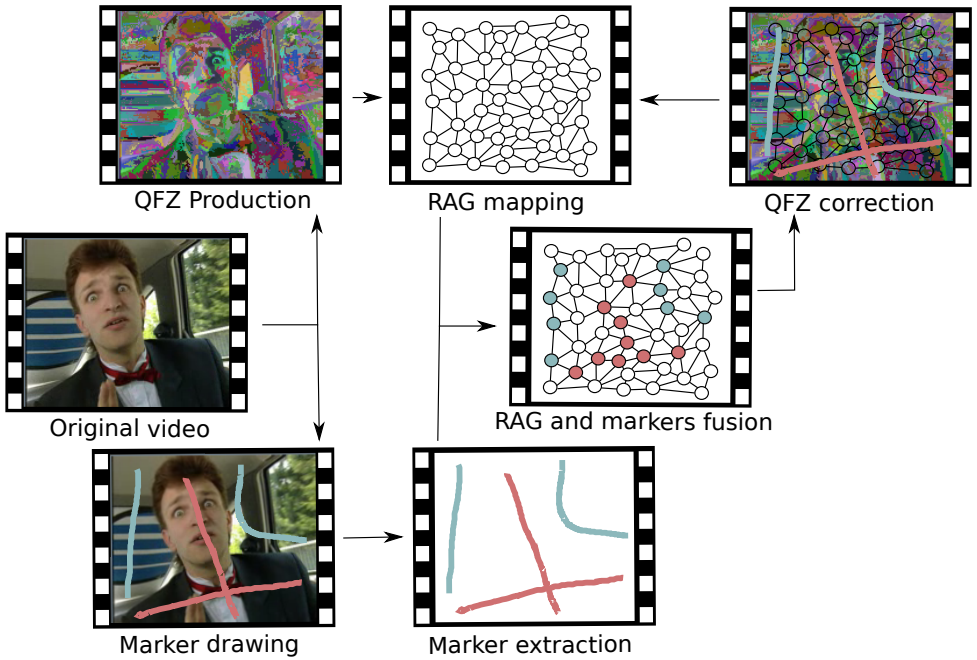
Interactive Segmentation: Framework



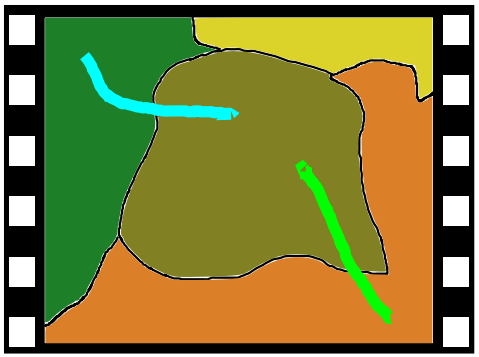
Interactive Segmentation: Framework

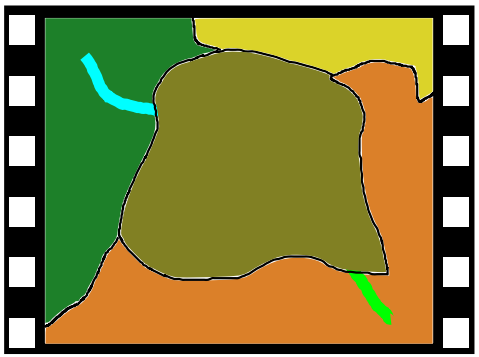


Interactive Segmentation: Framework

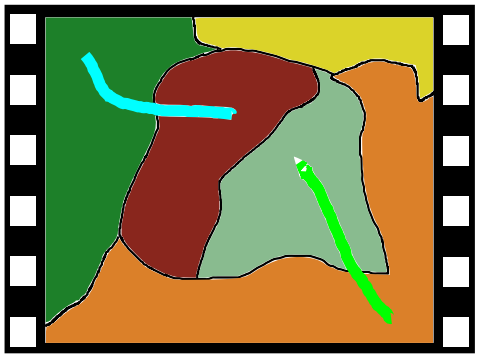


Interactive Segmentation: Framework



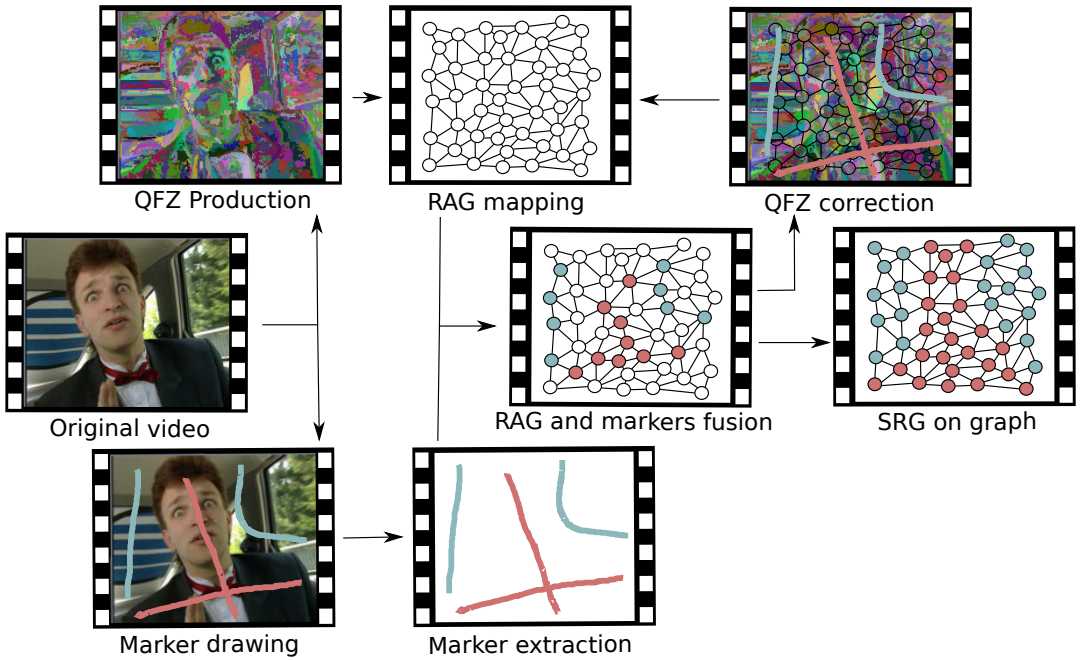


Solution 1: Edit markers

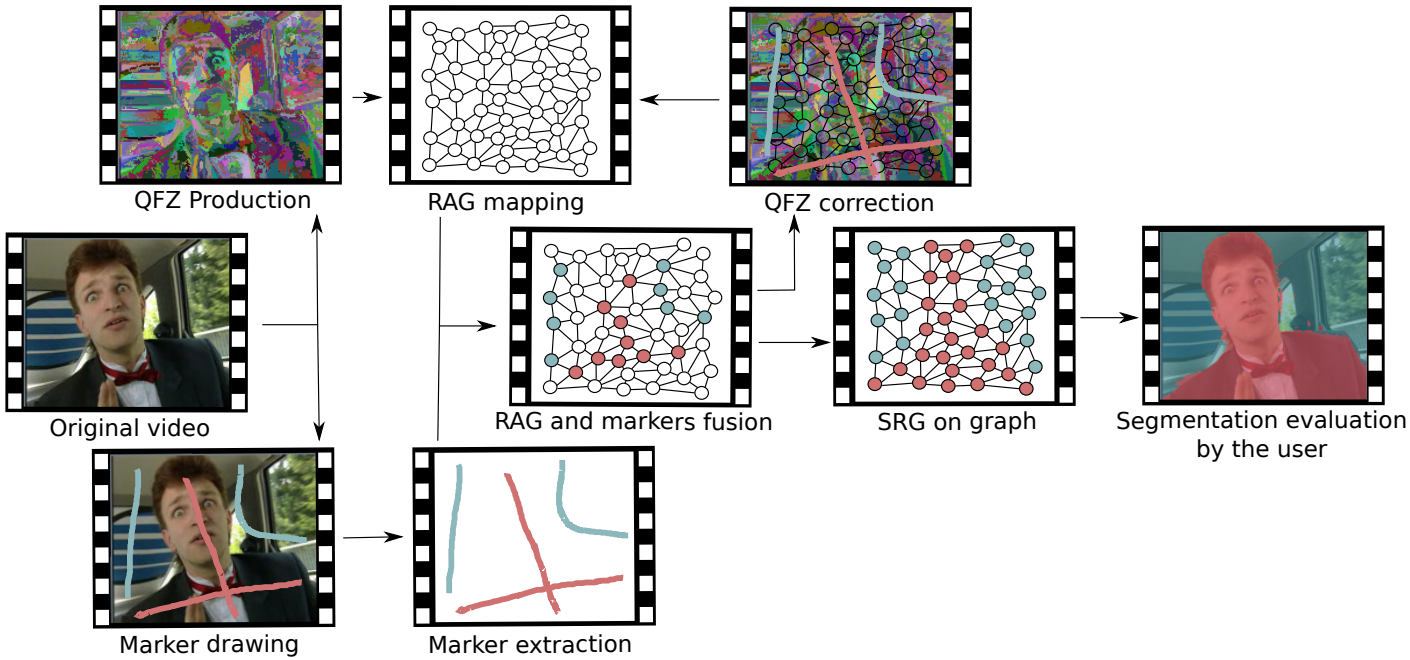


Solution 2 : Resegment QFZ according to the markers

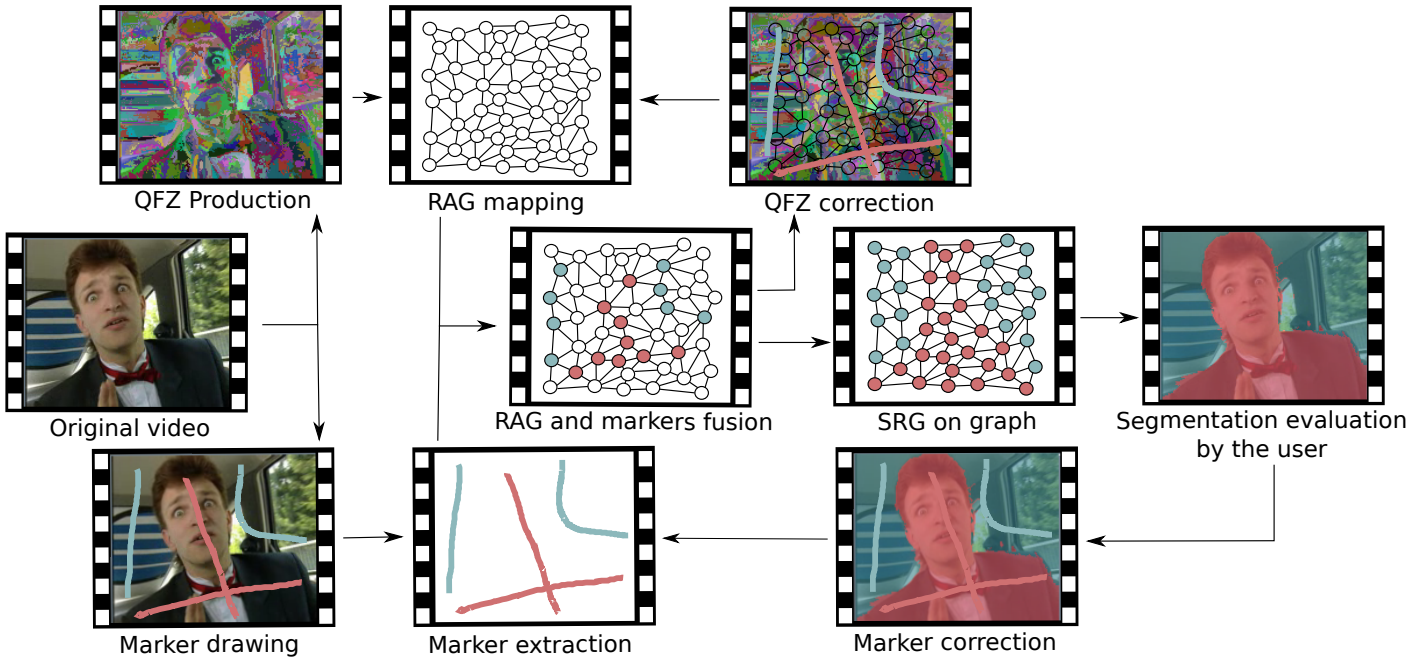
Interactive Segmentation: Framework



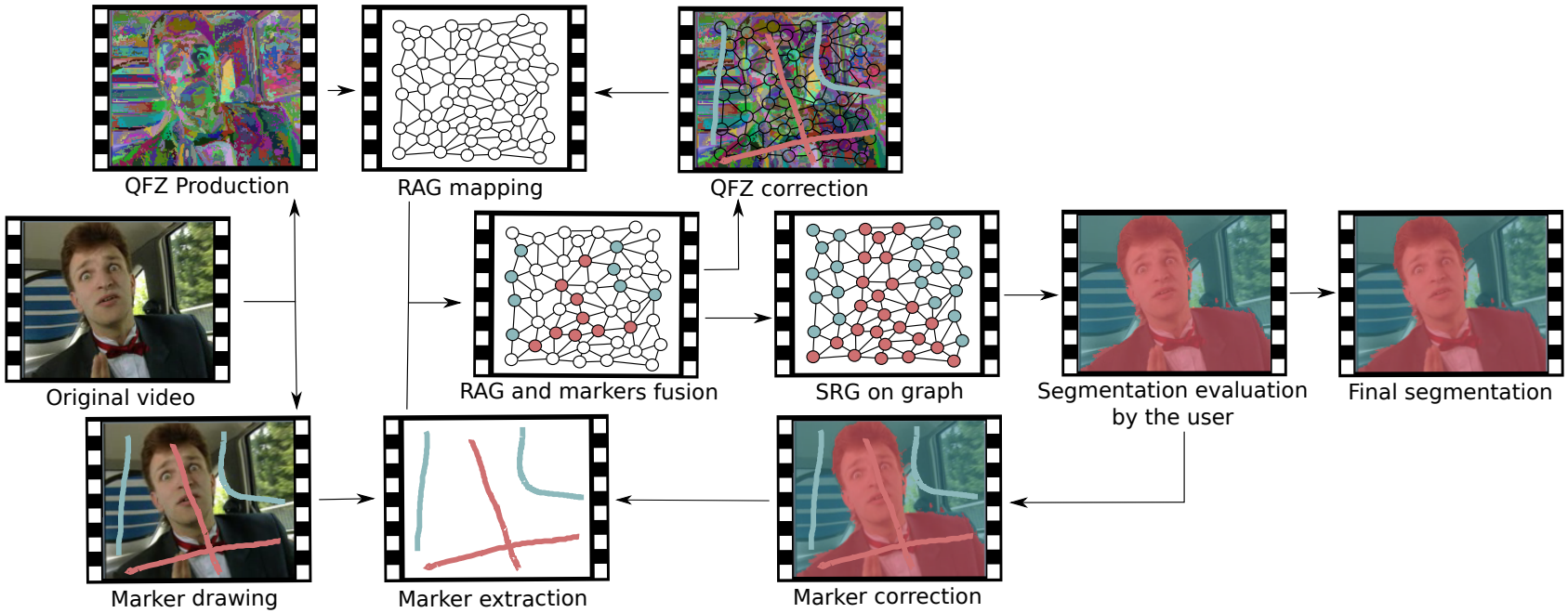
Interactive Segmentation: Framework



Interactive Segmentation: Framework



Interactive Segmentation: Framework



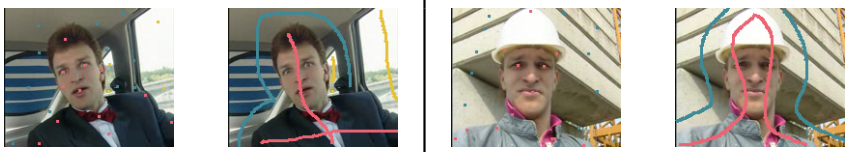
Computing time

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 *carphone* sequence (9 656 064 pixels, 176 x 144 on 381 frames).

Precision comparison

carphone Set 1 | *carphone* Set 2 | *foreman* Set 1 | *foreman* Set 2



Method	α, ω	*area**	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

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

- 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)

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

- 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)

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

- 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)

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

- 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)

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

- 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)

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

- 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)

j.weber@unistra.fr