Algorithm and VLSI Architecture for Real-time 1080p60 Video Retargeting

Pierre Greisen[1,2], Manuel Lang[1,2], Simon Heinzle[2], Aljosa Smolic[2]

1 ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

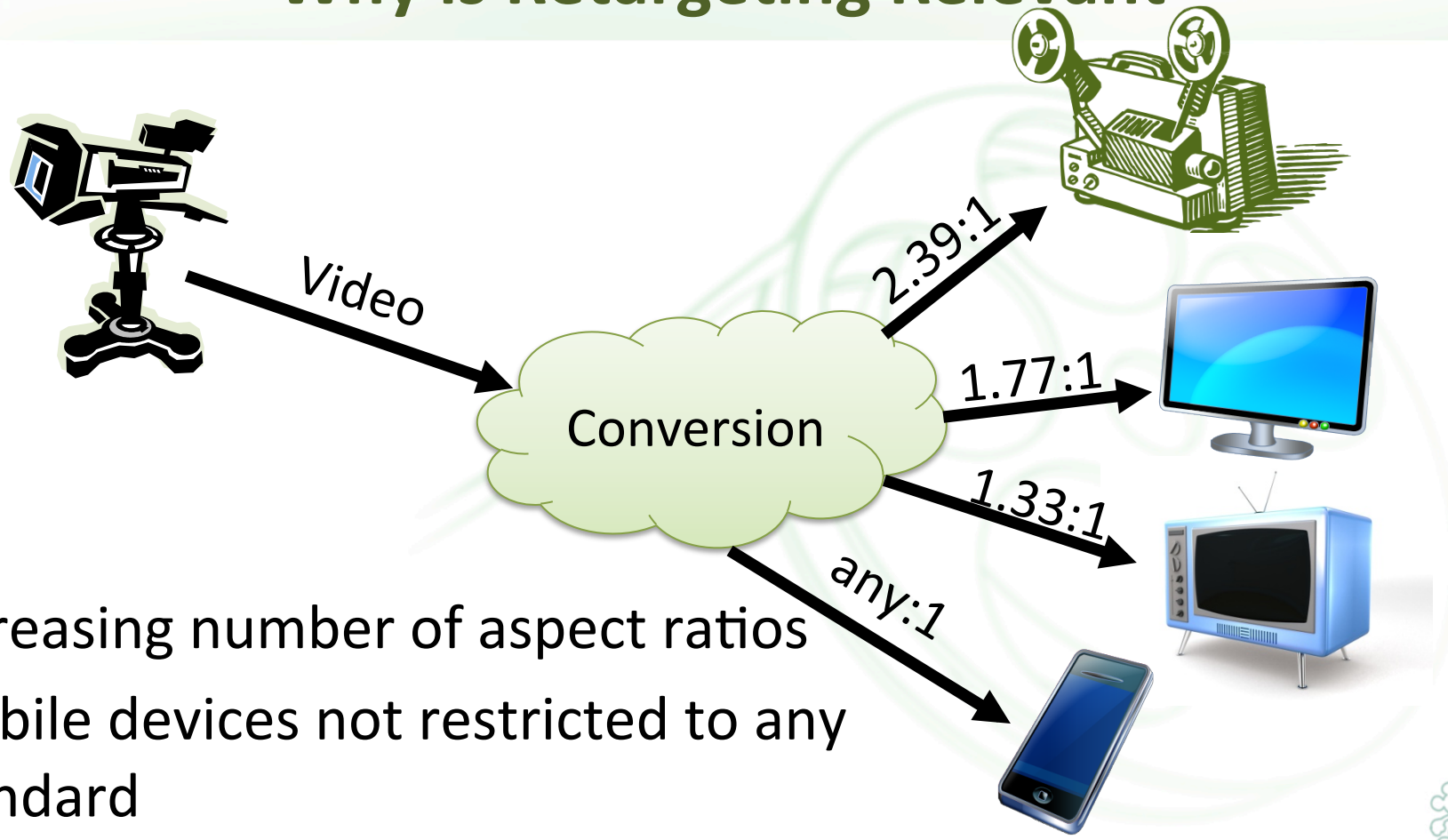
2 Disney Research, Zurich

What is Aspect Ratio Retargeting?

- ‘Clever’ change of aspect ratio
- Keep aspect ratio of important parts
- Distort / remove visually not important parts

[Kraehenbuehl 2009]

Why is Retargeting Relevant

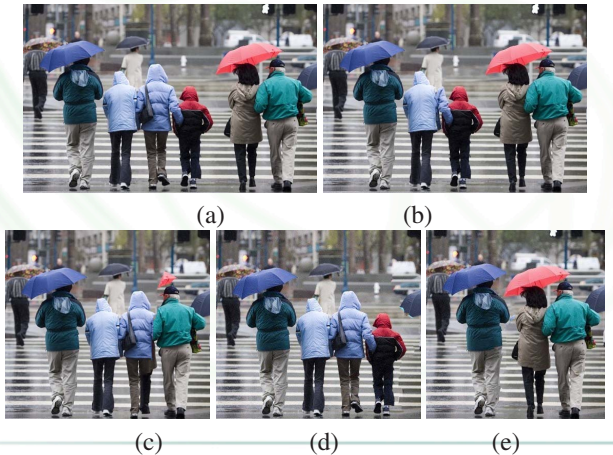


- Increasing number of aspect ratios
- Mobile devices not restricted to any standard

Discrete Methods

- Remove/Add Pixels
- Related work*:
  - Seam Carving [SIGGRAPH 07]
  - Improved Seam Carving [SIGGRAPH 08]
  - Shift-Map Image Editing [ICCV 09]

From: Avidan et.al., Seam Carving for Content-Aware Image Resizing

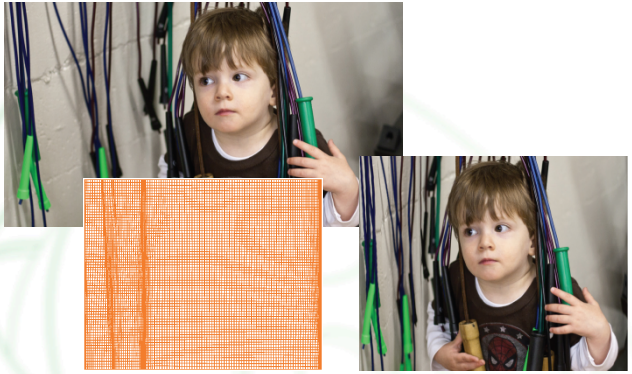


From: Pritch et. al., Shift-Map Image Editing

* non-exhaustive

Continuous methods

- Find continuous transformation
- Warp/deformation grid
- Related work*
  - Non-homogenous warping, ICCV 07
  - Streaming video, SIGGRAPH 09
  - Shrinkability Maps for Content-Aware Video Resizing, PG 09
  - Robust Image Retargeting via Axis-Aligned Deformation, EG 12



Generated with the Streaming Video approach [KRA09]

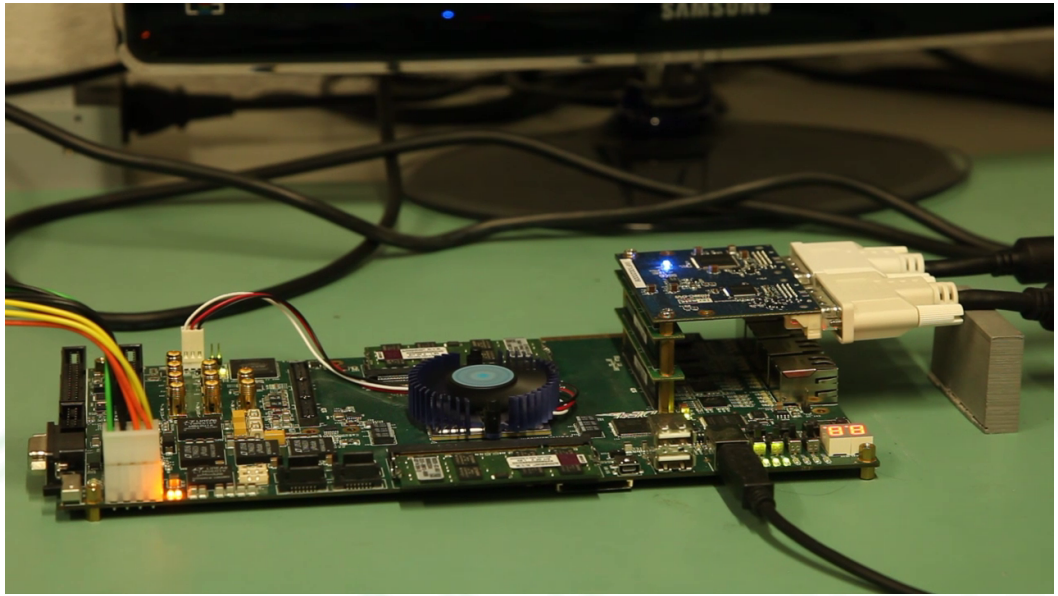
original saliency map

retargeting to 200% width using axis-aligned deformations

From: Robust Image Retargeting via Axis-Aligned Deformation

This Work

- Contributions:
  - *Video* retargeting
  - Efficient warp grid solver
  - First real-time implementation (FPGA)
  - 1080p60 at low HW resources

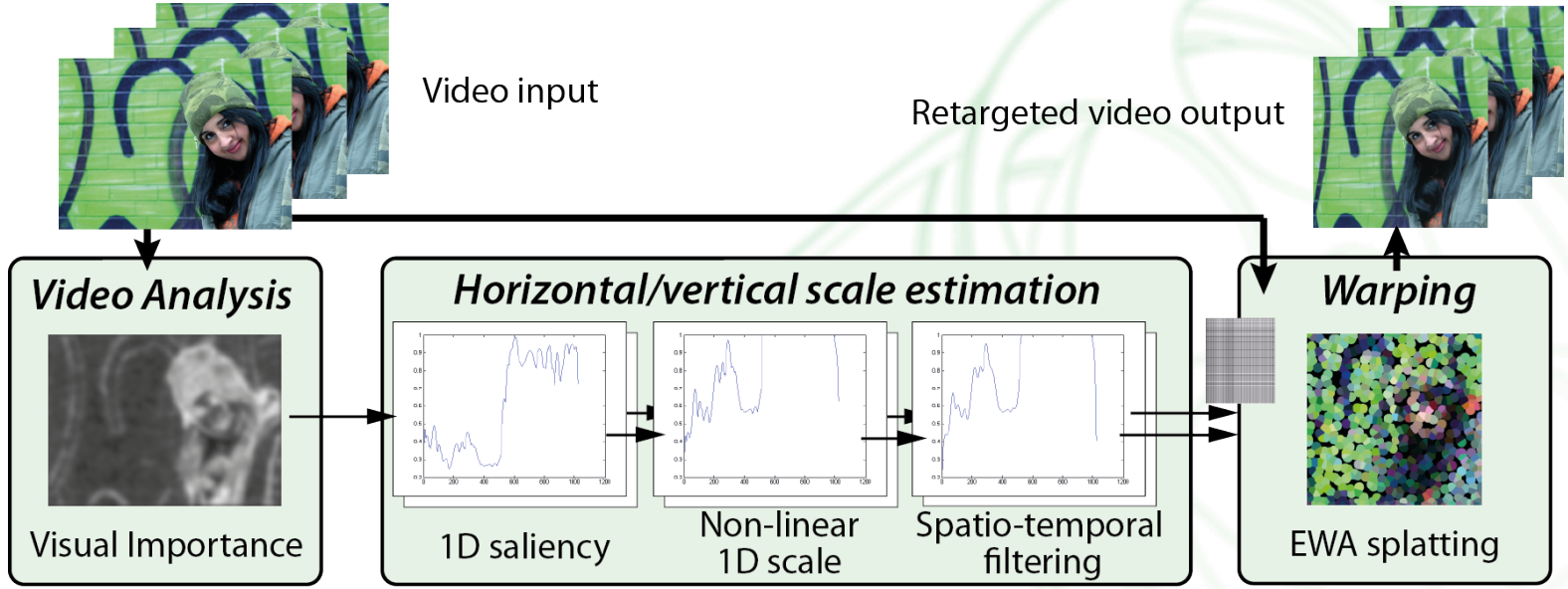


- Adaption of [KRA09] warping
  - [KRA09] among the best in RetargetME

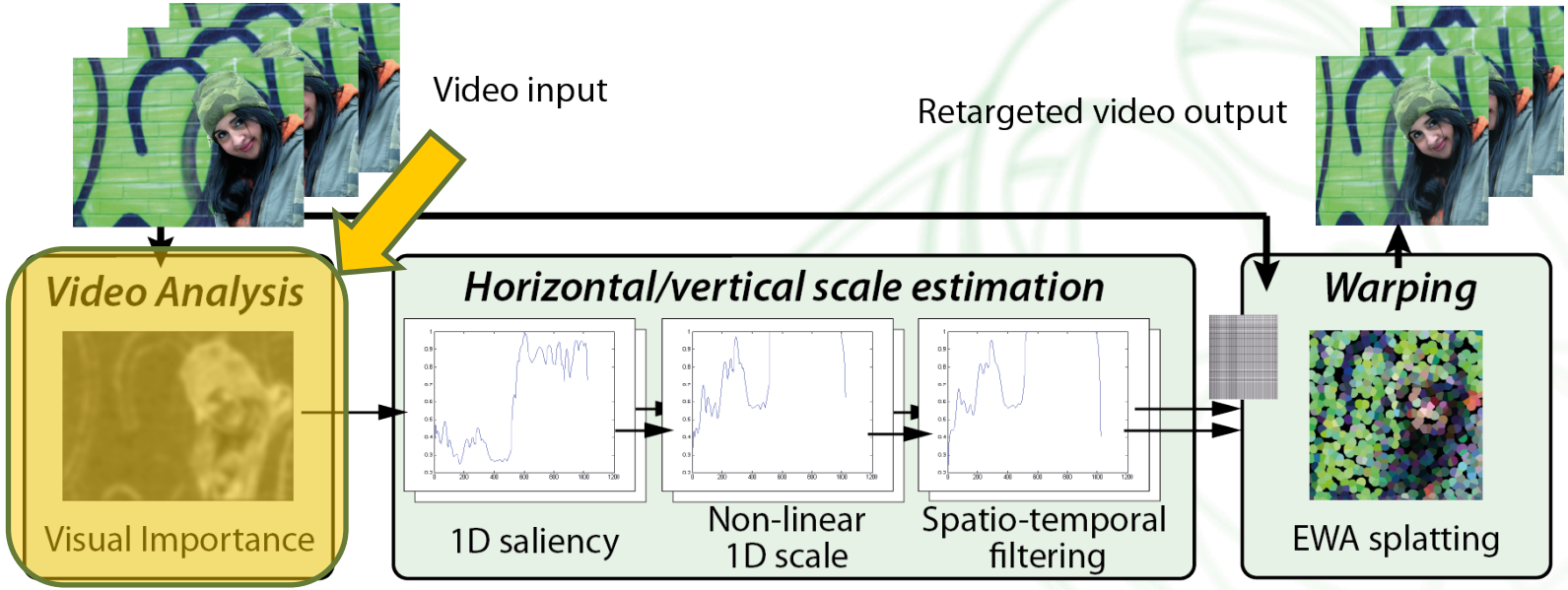
Outline

- Video Retargeting Algorithm
- Hardware Architecture (FPGA)
- Results & Limitations

Algorithm Overview

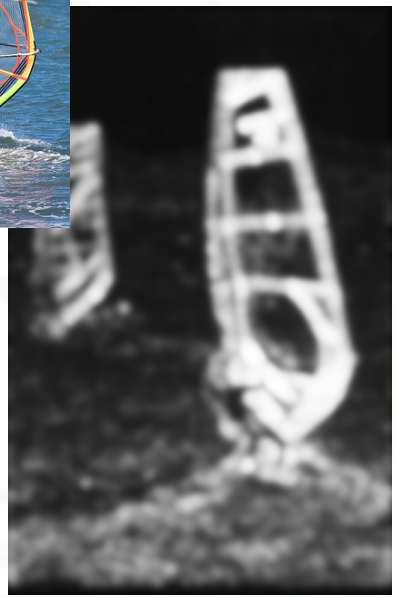


Algorithm Overview



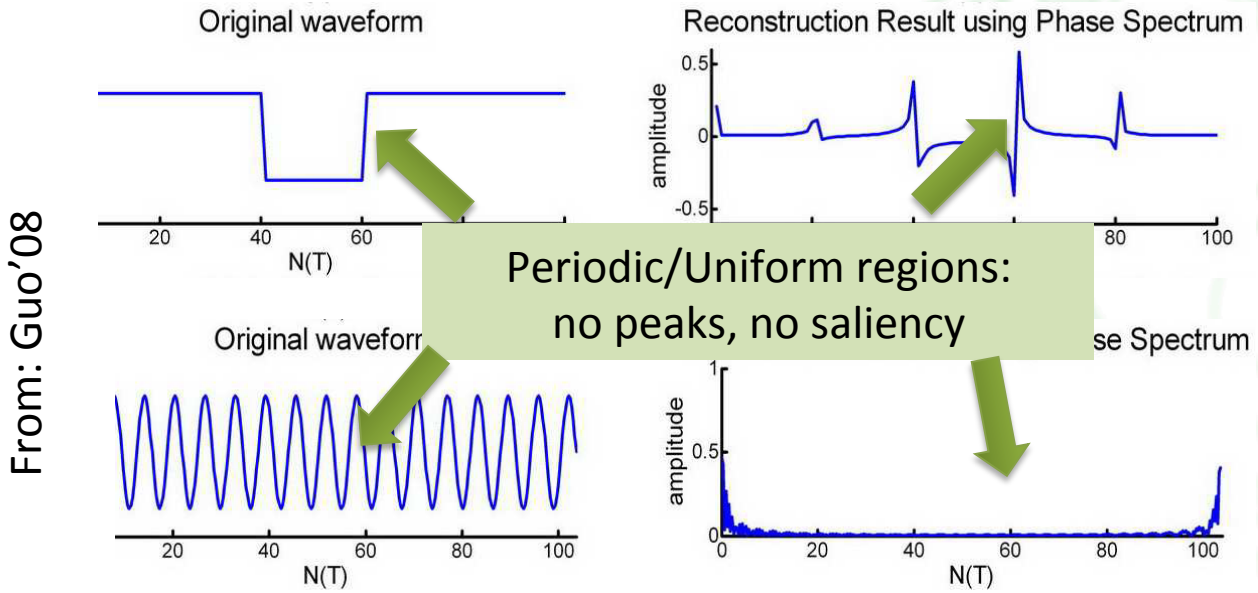
Visual Importance

- Saliency: where do people look, what is important in an image
- Video: moving objects
- No ‘silver bullet’ (yet?)
- Employed Algorithm:
  Guo 08: Spatio-temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform



Saliency from Phase Information

q(t) ⇨ Fourier Transform ⇨ Q(f)/|Q(f)| ⇨ Inv. Fourier Transform ⇨ q'(t)

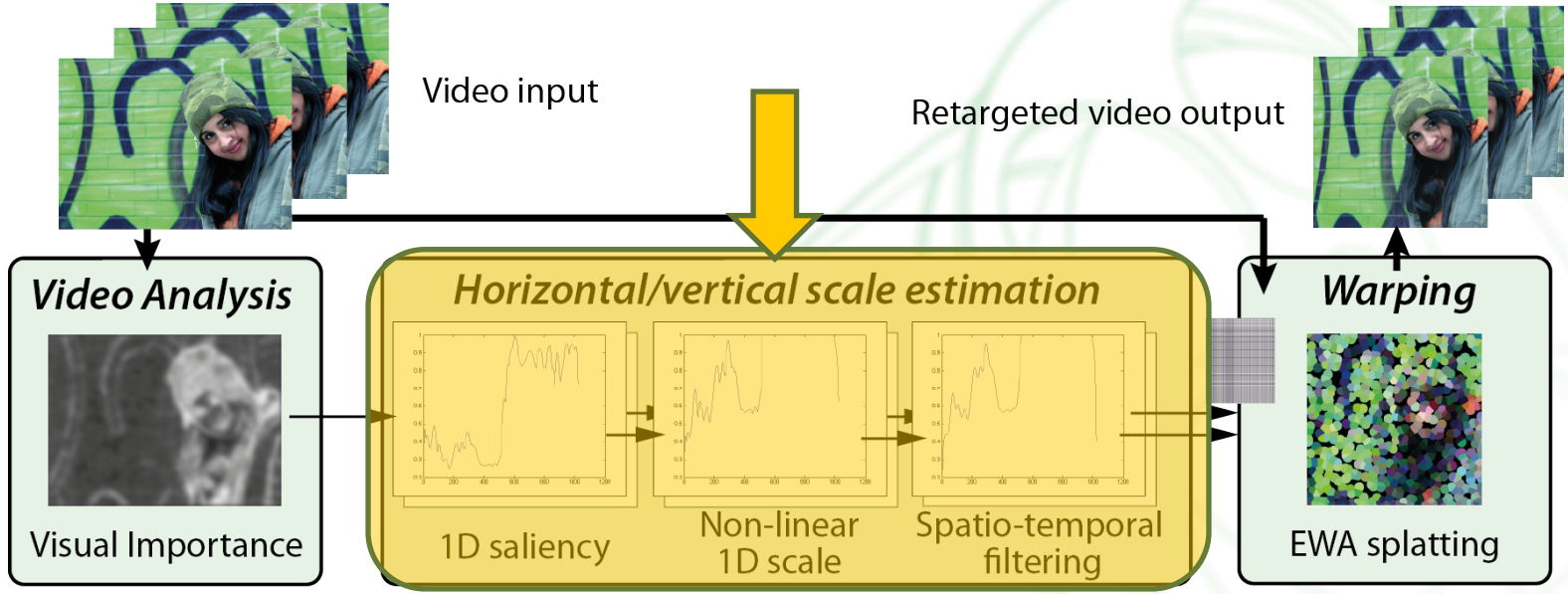


From: Guo'08

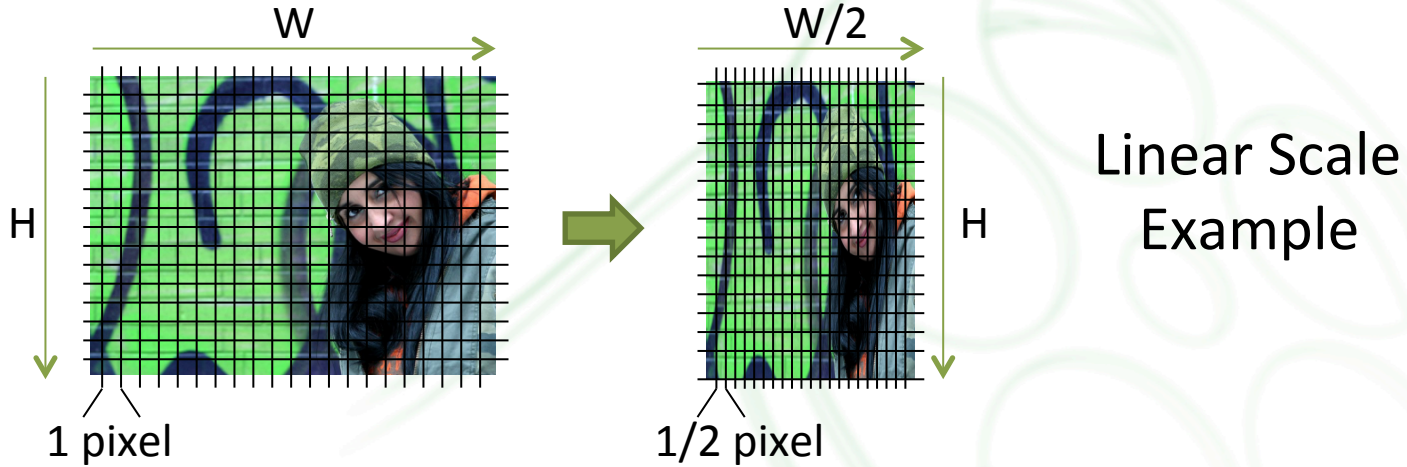
Real-time Spatio-temporal Saliency

- ‘Video phase’: 4 DOF (RGB + motion): Quaternion Fourier transform
- QFT = two 2D FTs
- Hardware efficiency:
  - Decompose image in blocks W x H/n
  - Saliency estimation per block
  - Blocks overlapping
  - Normalization

Algorithm Overview



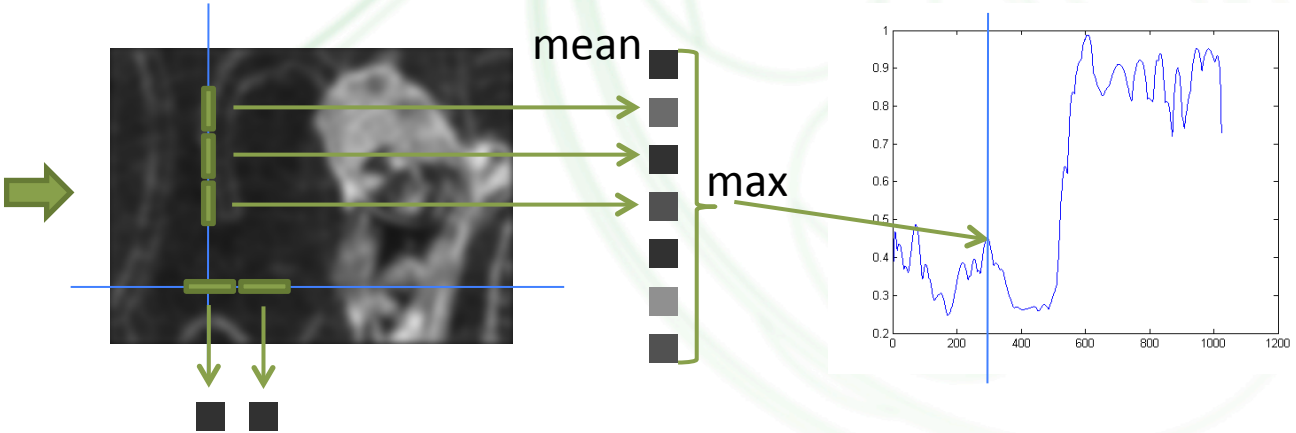
Deformation Grid Concept



- Per-pixel deformation value (in pixels)
- Cumulative sum = Position Grid

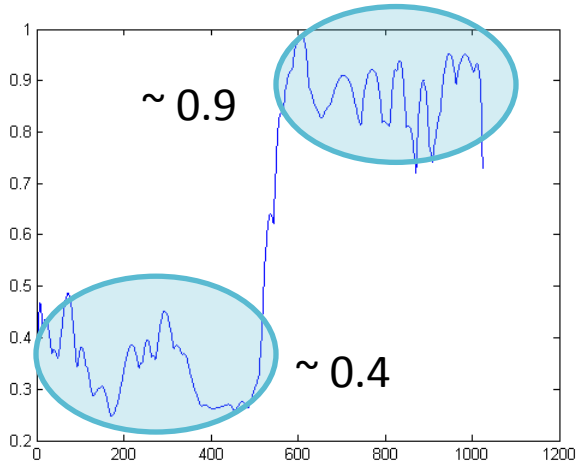
1D Saliency Profiles

- Project 2D saliency onto 1D axes
- Block-maximum as projection operator
- Detect small but salient objects



1D Profile as Local Resize Factor

- 1D Profile: Values between 0 and 1
- Profile value for local resizing/downscaling
- Magnification: mirror profile

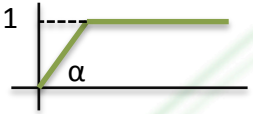


[W, H]

[1/2W, H]

How to scale?

- Retargeting constraint (W →W‘)
- Salient regions: resizing of 1
- Non-linear scaling



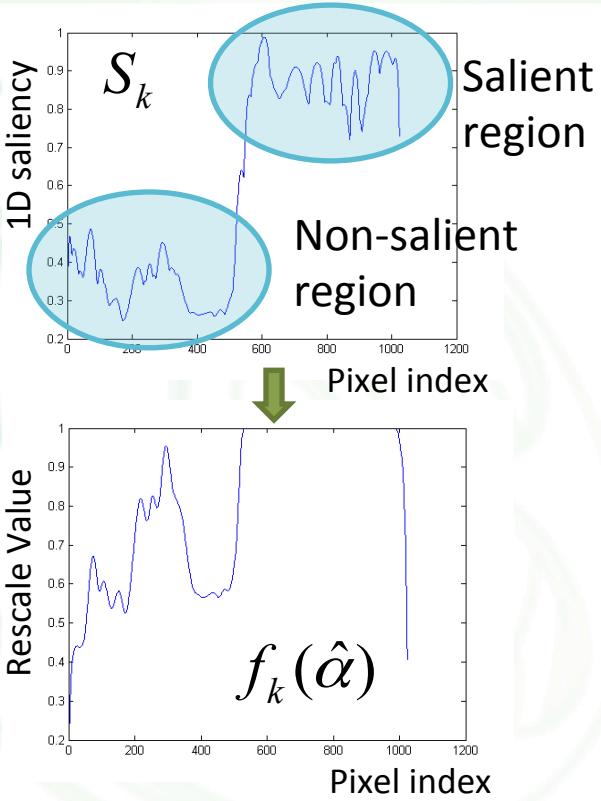
- Solve for α (binary tree search)

$$f_k(\alpha) = \min(\alpha S_k, 1), \quad s.t. \sum_{k=1}^{W} f_k(\alpha) = W'$$

rectilinear difference grid

1D projected saliency

retargeting constraint



Temporal Filtering of the Grid

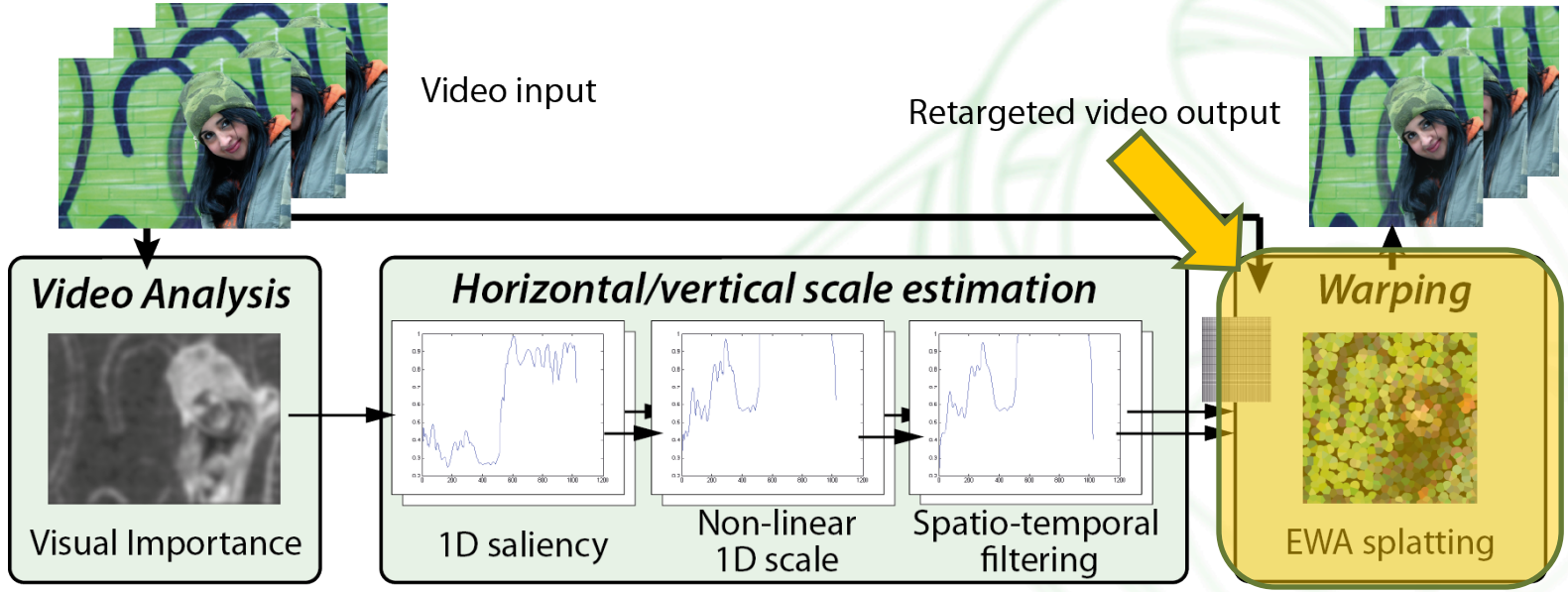
- Gaussian blur → latency
- IIR: y[k] = a y[k-1] + x[k] → a constant
- Determine ‘a’ based on scene motion

No Filtering

FIR (Gauss) only

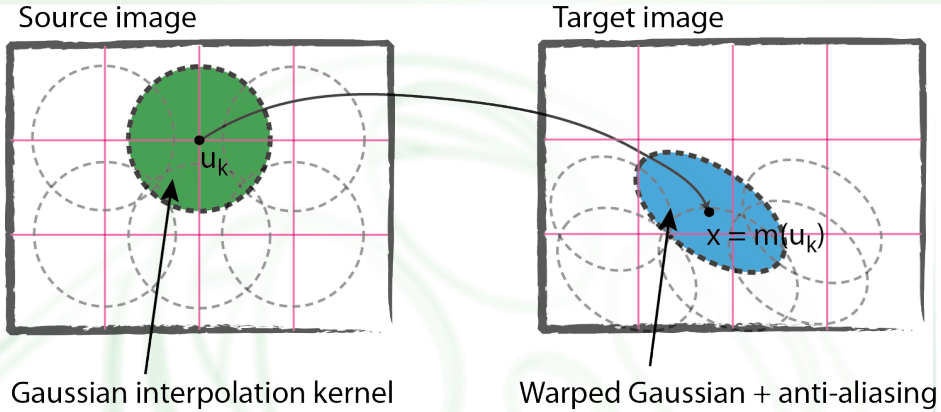
+ IIR

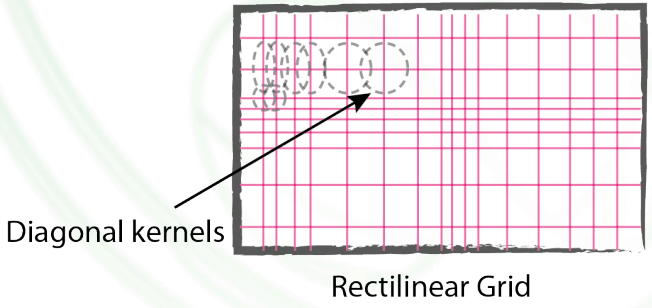
Algorithm Overview



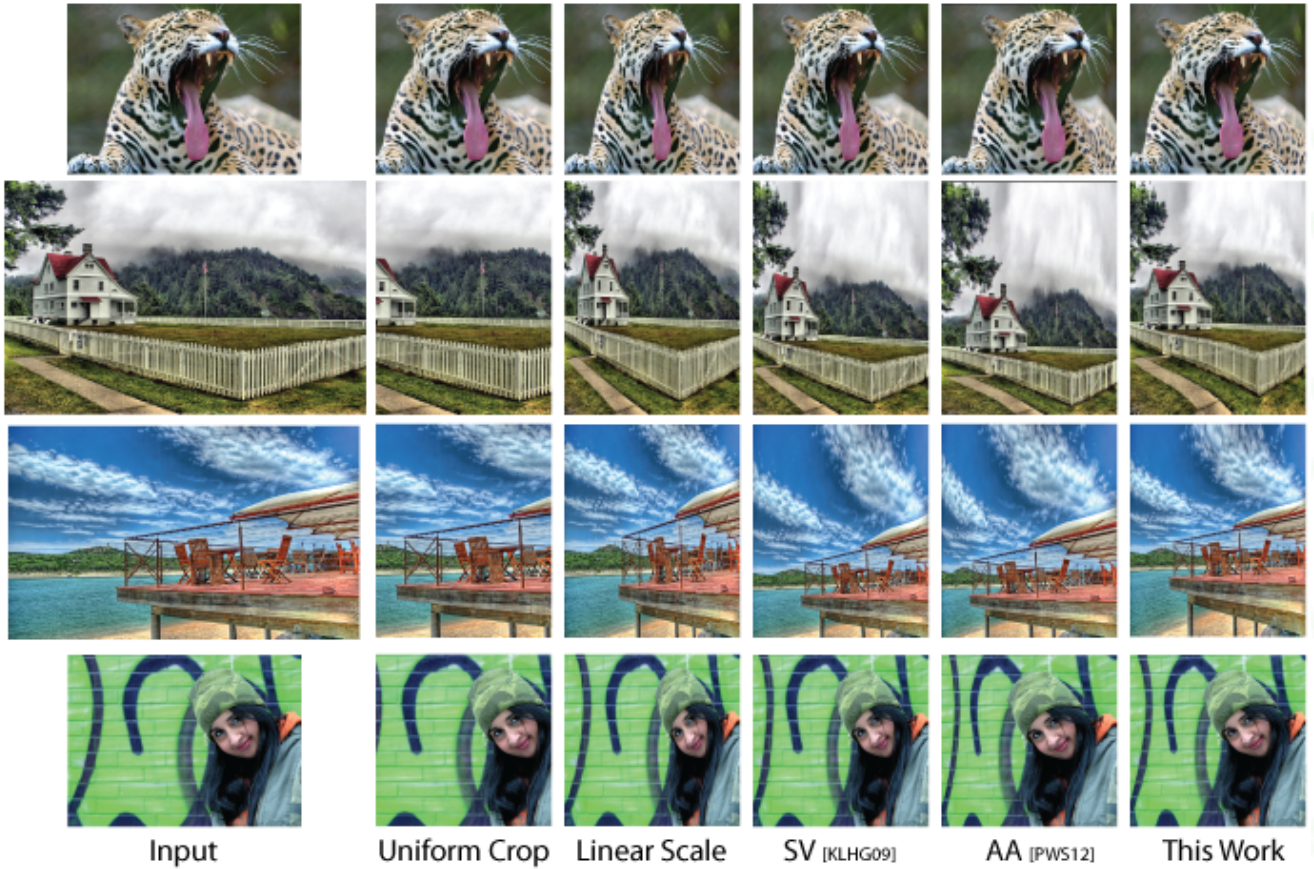
Rendering

- Spatially varying warping
  - EWA forward mapping
- Rectilinear grid:
  - Simplified EWA setup
  - Diagonal kernels (ops.)
  - Vertical deviations (buffer)
- Hardware-efficient and good quality (AA)





Algorithm Results



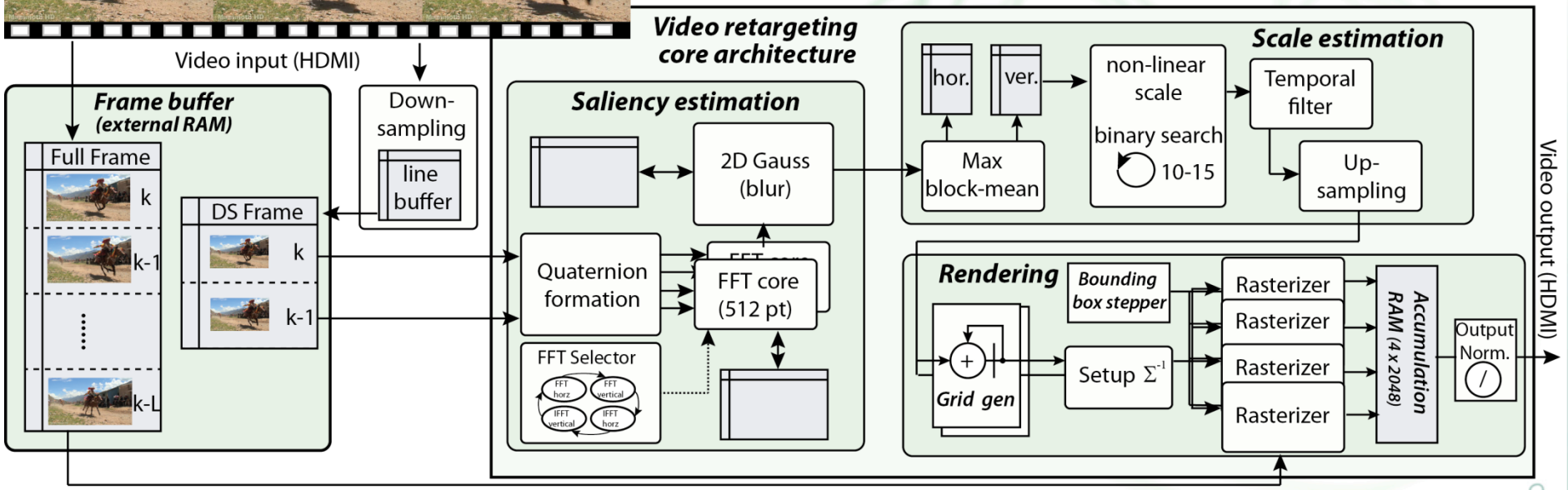
Images from RetargetMe Evaluation, Siggraph Asia 2010

Why Hardware?

- Video retargeting as fixed-function HW core in end-user devices
  (displays, mobiles, TVs, set-top boxes, …)
- Low-power @ high resolution
- Real-time capability of HD video retargeting
- FPGA HW Architecture:
  - VHDL
  - Pipelined stream processing
  - 1080p60 performance

Hardware Architecture

- FPGA board with HDMI, external frame-buffer



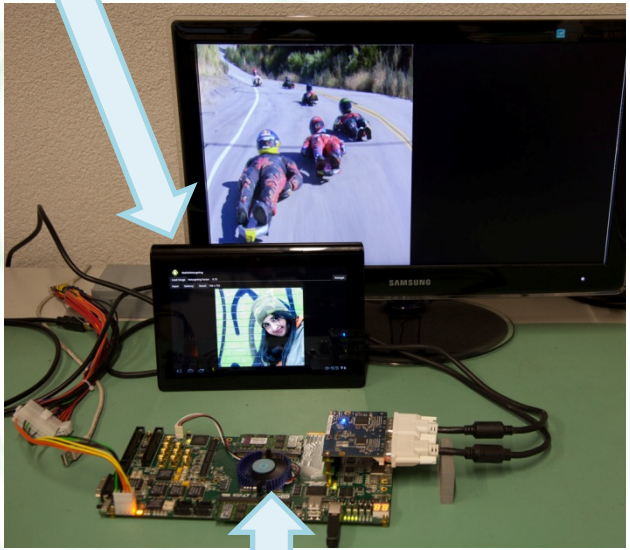
Implementation Results

FPGA Resources	Logic (LUTs)	Register bits	Block RAM	DSP slices
Saliency	8690	13571	847K	115
Grid Gen.	993	562	87K	12
Rendering	4071	2731	483K	80
Total Core	13762	16864	1416K	207
Cyclone IV	**12%**	**15%**	**36%**	**78%**

Performance	FPGA	Tegra II
	1080p60 @ 130 MHz	1080p2

ALTERA Cyclone IV EP4CE115 = low-end 60nm FPGA

Tegra II tablet



ALTERA FPGA board (Terasic)

Video Results

Conclusion

- Limitations
  - Temporal artifacts
  - Saliency estimation quality
  - No integration into CE products, yet
- But …
  - Motion estimation
  - Temporal stable saliency (recent work)
  - Meta-information
  - Compression
  - Generalization/extension to more video/graphics applications (STEREO/MV)