Technical University of Lodz
Institute of Electronics

Selecting a Motion Estimation Method for a Model of Deformable Rings

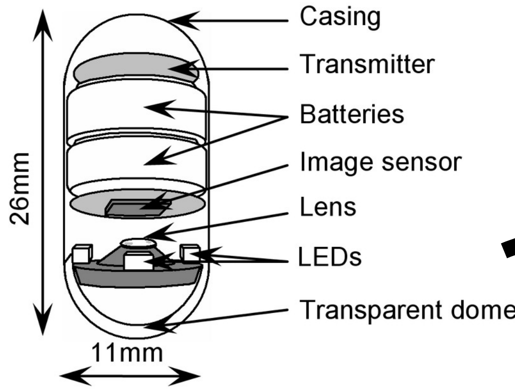
Piotr M. Szczypiński

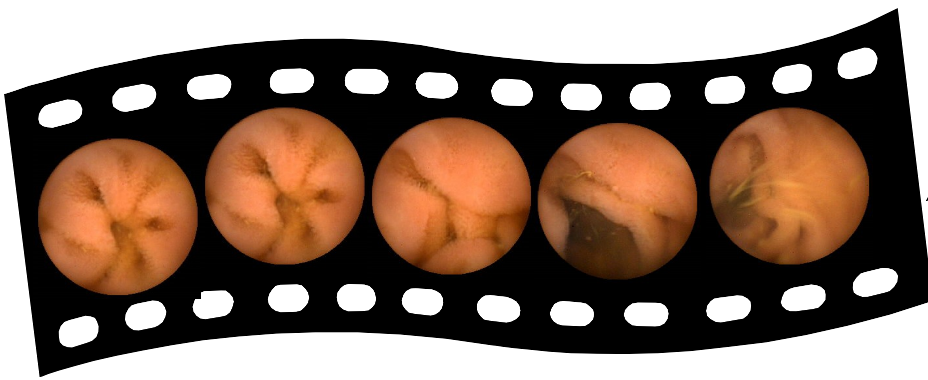
ICSES'06

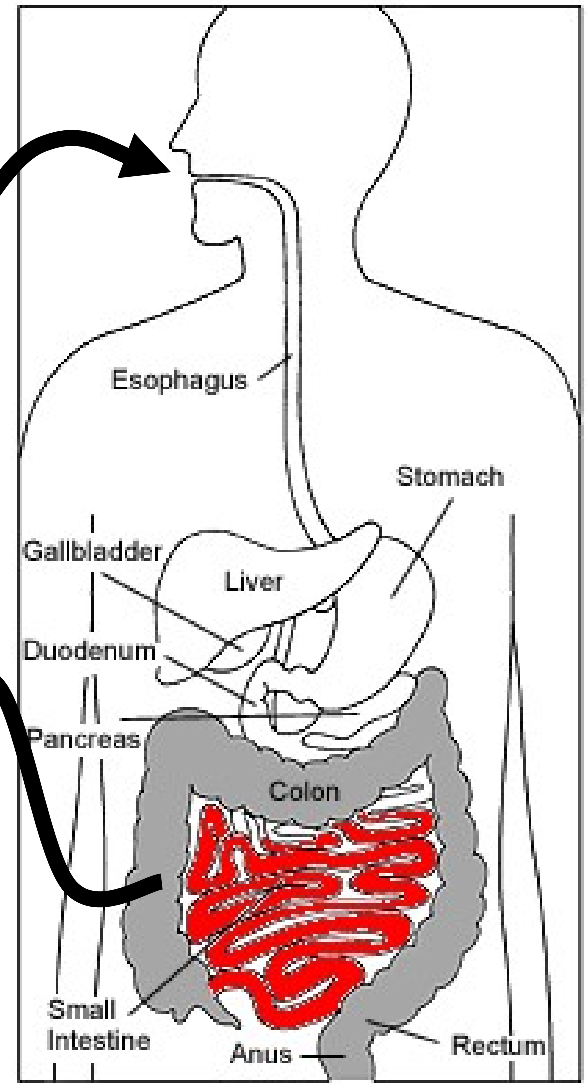
Scope of the presentation

- What is Wireless Capsule Endoscopy (WCE)
- How the Model of Deformable Rings (MDR) aids in WCE interpretation
- How the MDR works
- Variants of motion estimation methods for MDR
- Test results, comparisons and conclusions

Wireless Capsule Endoscopy







Two frames per second, 8-hour video, 256 × 256 pixels,
circular, 140° field of view, about 240 pixels in diameter

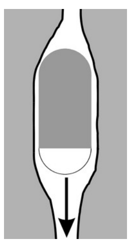
Wireless Capsule Endoscopy

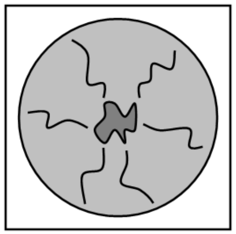
The video interpretation involves viewing the video and searching for bleedings, erosions, ulcers, polyps and narrow sections of the bowel or any other abnormal-looking entities.

Problems:
Interpretation takes considerable amount of time, usually more than an hour per recording, is performed by a highly trained clinician, it is a tedious task.

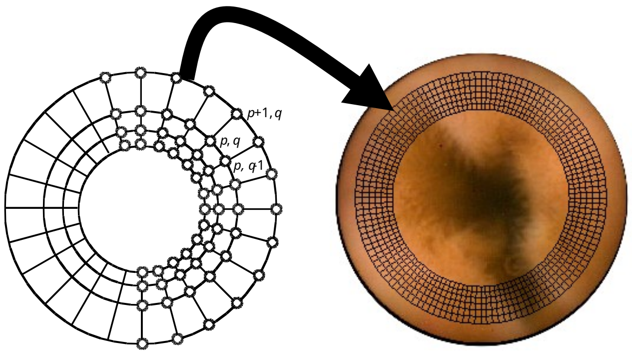
Model of Deformable Rings

<u>What we assume:</u>
Since the shape of the capsule is elongated and the GI tract is akin to a collapsed tube, most of the time the wireless capsule endoscope aligns in a direction parallel to the GI tract.
Most of the video frames contain images of the GI tract walls, which converge in perspective at the point located near the center of an image.
As the capsule passes through, portions of the GI tract image shift outward (while capsule moves forward) or toward the center of an image (while capsule moves backward).





Model of Deformable Rings

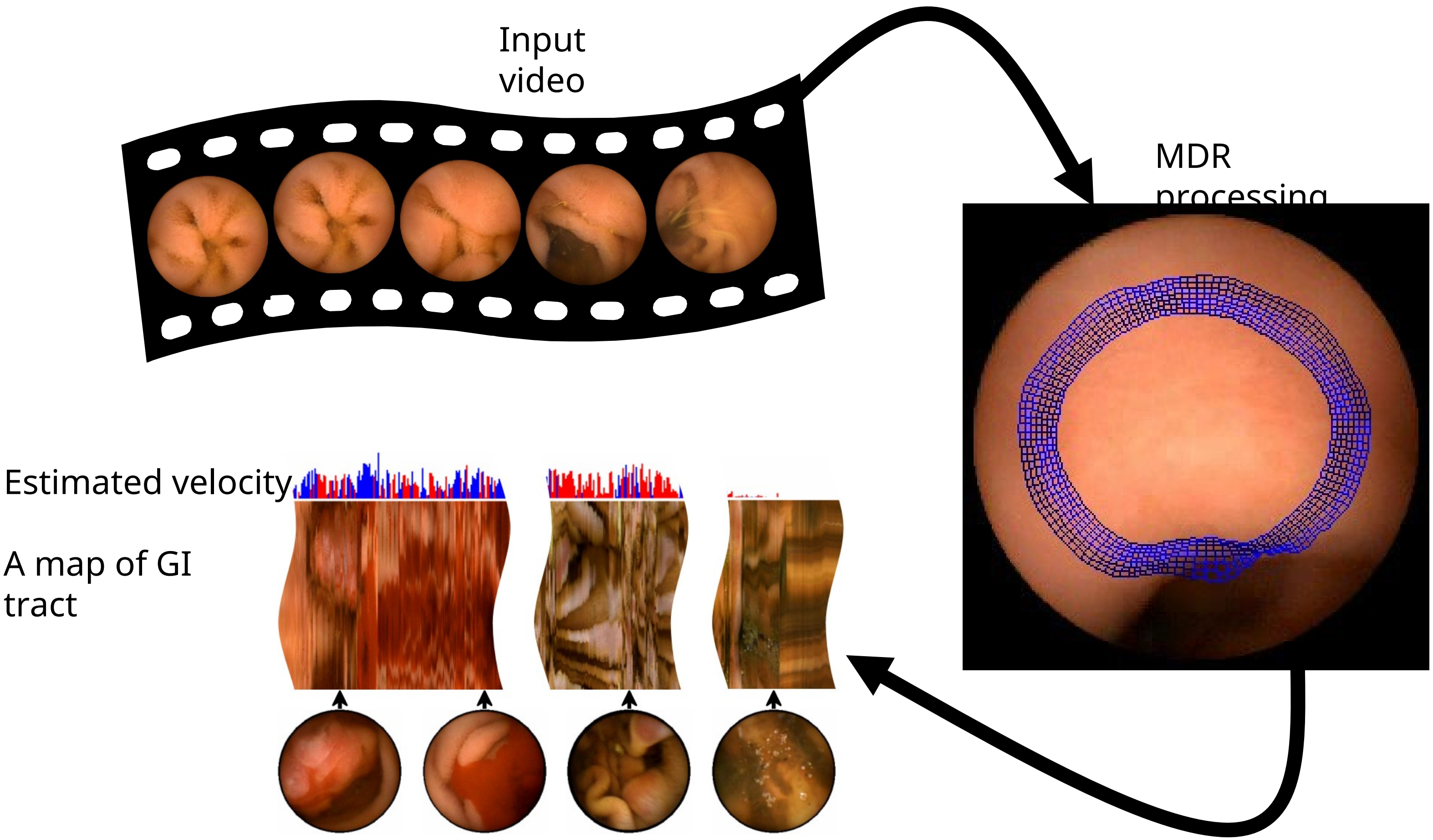


<u>The model:</u>
The MDR is composed of interconnected nodes, which form concentric rings having a center near the center of the image frame.
The aim of the MDR is to track motion of digestive system walls within the video.

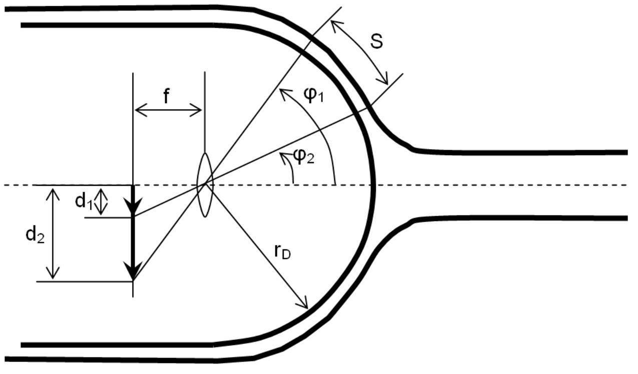
P.M. Szczypinski, P.V.J. Sriram, R.D. Sriram, and D. Reddy, “Computerized Image Analysis of Wireless Capsule Endoscopy Videos Using a Dedicated Web-like Model of Deformable Rings – A feasibility Study,” Endoscopy 2004; 36 (Suppl I) A76

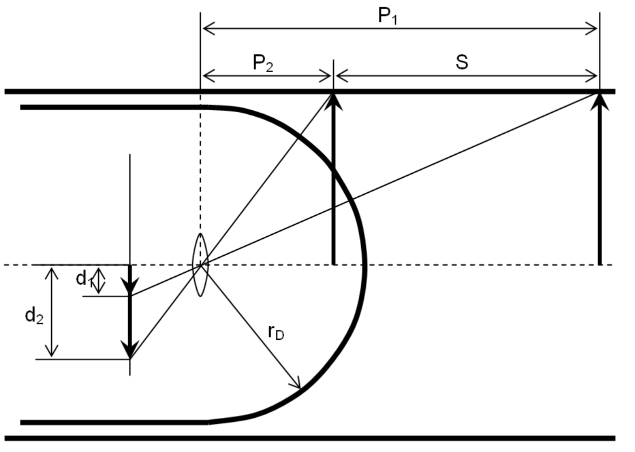
P.M. Szczypinski, P.V.J. Sriram, R.D. Sriram, and D.N. Reddy, “Model of Deformable Rings for Aiding the Wireless Capsule Endoscopy Video Interpretation and Reporting,” Computer Vision and Graphics, International Conference, ICCVG’04, 2006 Springer: pp.167-172

Model of Deformable Rings



Velocity Estimation

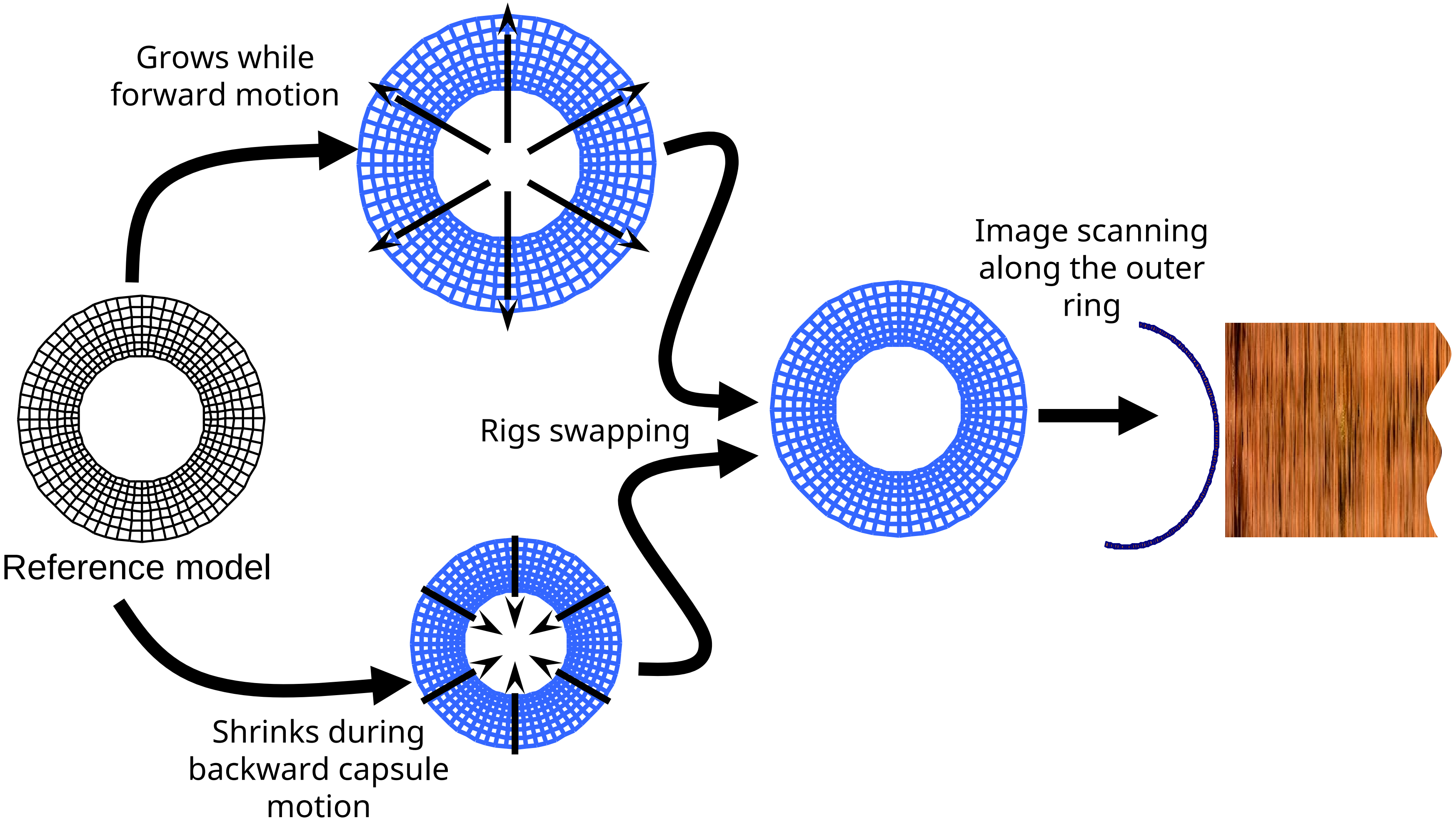




$$V_D = \frac{r_D}{\Delta t}\left(\tan^{-1}\left(\frac{d_1}{f}\right) - \tan^{-1}\left(\frac{d_2}{f}\right)\right)$$

$$V_P = \frac{r_D f}{\Delta t}\left(\frac{1}{d_2} - \frac{1}{d_1}\right)$$

MDR Map



Grows while forward motion
Reference model
Rigs swapping
Image scanning along the outer ring
Shrinks during backward capsule motion

MDR Motion

$$E_{MDR} = \rho E_i + \xi E_e$$

Models energy
to be
minimized

Internal
tensions

Image
influence
component

ρ, ξ - parameters

The main goal was:
How to design the image
influence energy for the MDR?

Image dissimilarity functions

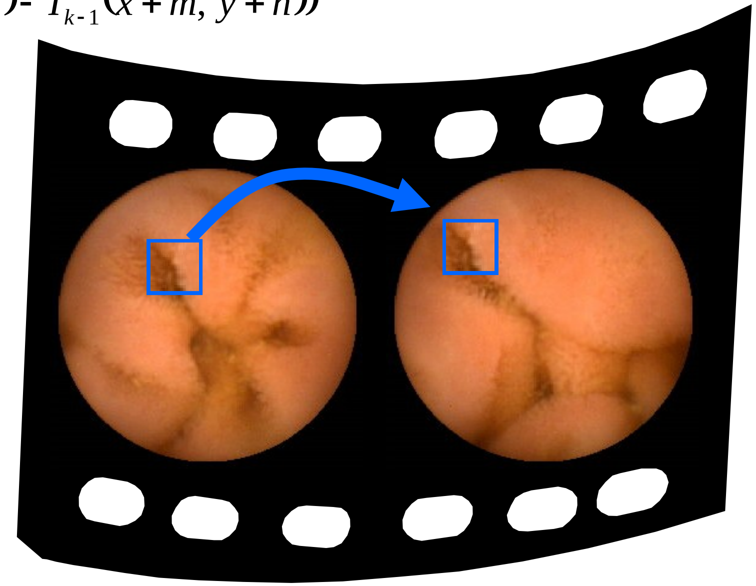
mean absolute difference function:

$$MAD(\Delta x, \Delta y) = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \left| I_k(x + \Delta x + m, x + \Delta y + n) - I_{k-1}(x + m, y + n) \right|$$

mean square error function:

$$MSE(\Delta x, \Delta y) = \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \left(I_k(x + \Delta x + m, x + \Delta y + n) - I_{k-1}(x + m, y + n) \right)^2$$

I_k – current frame
I_{k-1} – previous frame of a video sequence
$N \times N$ – block size
Δx, Δy – vector of motion



Energy minimization procedure

Two methods for energy minimization were tested:

Full Search (FS) method

- the location for minimum value of dissimilarity function is computed individually for each node
- to minimize E_e component, node is pushed toward the location, with intensity proportional to the distance from the location

Gradient Based (GB) method

- the gradient of dissimilarity measure is computed at current location of each node
- to minimize E_e component, node is pushed according to the reversed gradient vector of the dissimilarity function

Variants of motion estimation methods

Selection of dissimilarity function:

- Mean absolute difference or mean square error function
- Size (*N*) of a block for matching

Energy minimization:

- Full Search or Gradient Based method
- Ratio between ξ and ρ parameters

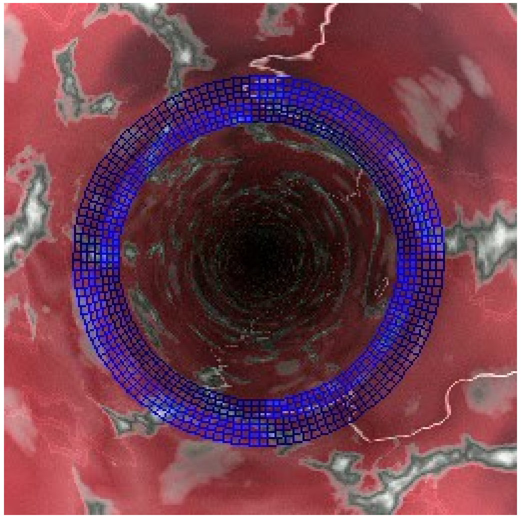
Video data for experiments

The MDR was tested on:

- artificially generated videos
- WCE video

The artificial videos demonstrate:

- known kind of motion
- within a rigid, texturized pipe



Example of MDR processing the video showing a constant forward movement with a constant rotation

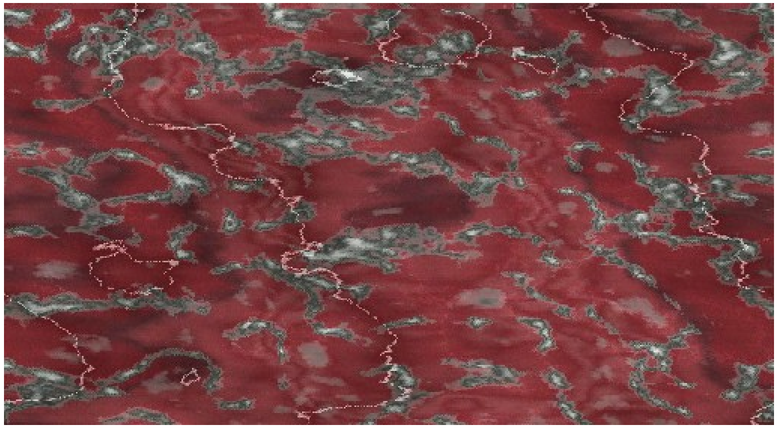
Results

FS vs. GB and ξ parameter

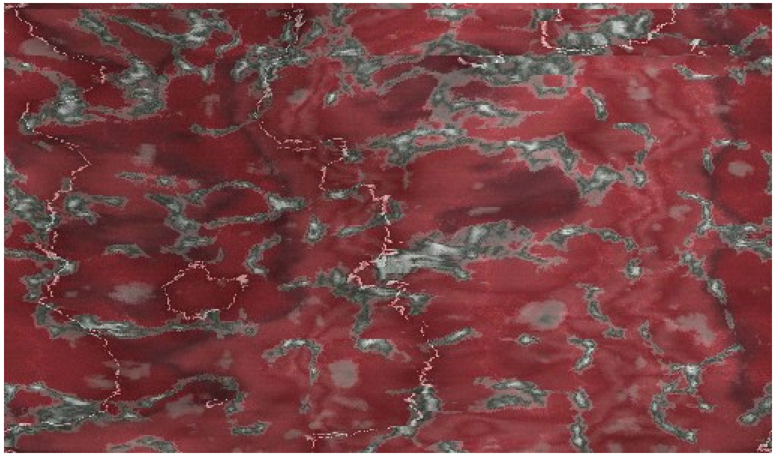
	ξ	V_P [mm/s]	$\frac{S(V_P)}{V_P}$	V_D [mm/s]	$\frac{S(V_D)}{V_D}$	Angle [deg.]	Defor-mation	Time [ms]
FS-MAD N=5	0.80	0.161	0.064	0.125	0.064	283.4	0.173	291
	0.40	0.163	0.065	0.127	0.065	284.9	0.140	291
	0.20	0.167	0.056	0.130	0.056	286.7	0.109	304
	0.10	0.164	0.053	0.128	0.053	263.9	0.072	299
	0.05	0.152	0.057	0.119	0.057	216.3	0.036	292
	0.05	0.124	0.087	0.096	0.088	139.3	0.022	231
GB-MAD N=5	0.40	0.177	0.100	0.138	0.101	359.7	0.541	153
	0.20	0.181	0.115	0.141	0.116	354.2	0.153	155
	0.10	0.181	0.102	0.141	0.102	353.9	0.044	155
	0.05	0.184	0.043	0.143	0.043	343.0	0.014	156
	0.02	0.157	0.094	0.123	0.094	178.0	0.003	153
Expected		0.2	0	-	0	360	0	min.

Results
FS vs. GB

FS, MAD, N=5



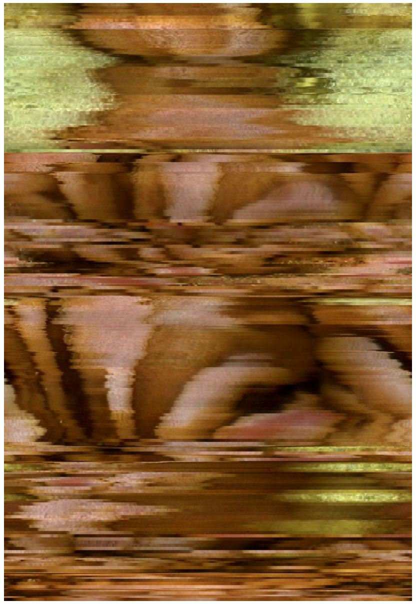
GB, MAD, N=5



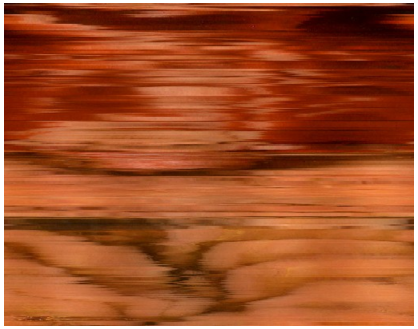
Results
FS vs. GB

FS, MAD, N=3

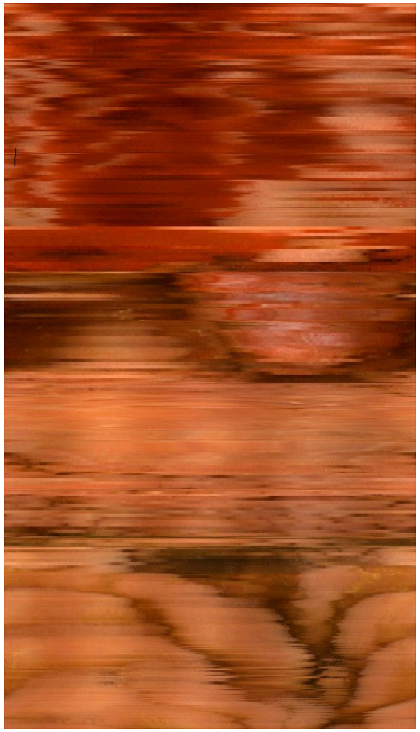
GB, MAD, N=3



FS, MAD, N=3



GB, MAD, N=3



Results

N=1 vs. N=5 and MAD vs. MSE

	ξ	V_P [mm/s]	$\frac{S(V_P)}{V_P}$	V_D [mm/s]	$\frac{S(V_D)}{V_D}$	Angle [deg.]	Defor-mation	Time [ms]
	0.40	0.177	0.100	0.138	0.101	359.7	0.541	153
	0.20	0.181	0.115	0.141	0.116	354.2	0.153	155
	0.10	0.181	0.102	0.141	0.102	353.9	0.044	155
	0.05	0.184	0.043	0.143	0.043	343.0	0.014	156
GB-MAD N=5	0.02	0.157	0.094	0.123	0.094	178.0	0.003	153
	0.40	0.180	0.178	0.140	0.178	354.6	1.858	22
	0.20	0.181	0.120	0.141	0.120	355.1	0.481	22
	0.10	0.185	0.073	0.144	0.073	355.9	0.124	23
	0.05	0.184	0.060	0.143	0.060	336.5	0.034	22
GB-MAD N=1	0.02	0.155	0.113	0.121	0.114	168.6	0.011	22
	0.80	0.173	0.405	0.135	0.400	359.1	0.761	123
	0.40	0.186	0.131	0.145	0.132	354.1	0.241	123
	0.20	0.188	0.059	0.146	0.059	356.3	0.072	123
	0.10	0.184	0.054	0.144	0.054	336.6	0.019	124
GB-MSE N=5	0.05	0.159	0.119	0.123	0.119	184.5	0.006	123

Conclusions

In MDR:

- The Full Search method, the best in video coding, is inferior to Gradient Based method
- There is no significant difference between use of MSE or MAD dissimilarity functions
- Bigger sizes of blocks for matching does not contribute significantly to the accuracy of motion tracking
- The higher the ξ parameter the better is motion tracking but at the same time the higher undesired deformation

Suggestions are to use GB method with MAD function,

and N = 1 or N = 3.