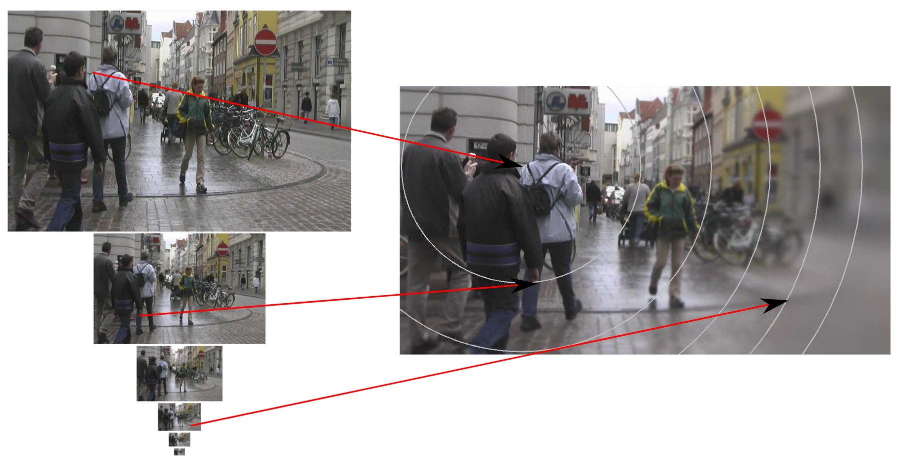
Real-time gaze-contingent spatio-temporal filtering for gaze guidance

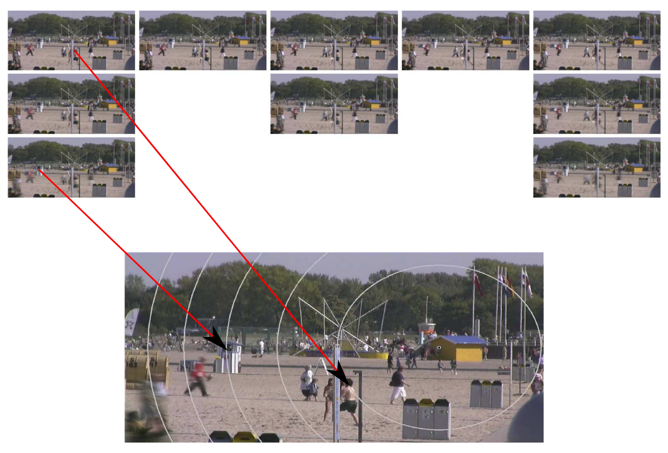
Michael Dorr, Sönke Ludwig, and Erhardt Barth

Institute for Neuro- and Bioinformatics, University of Lübeck, http://www.inb.uni-luebeck.de
{dorr|ludwig|barth}@inb.uni-luebeck.de

'Foveation'

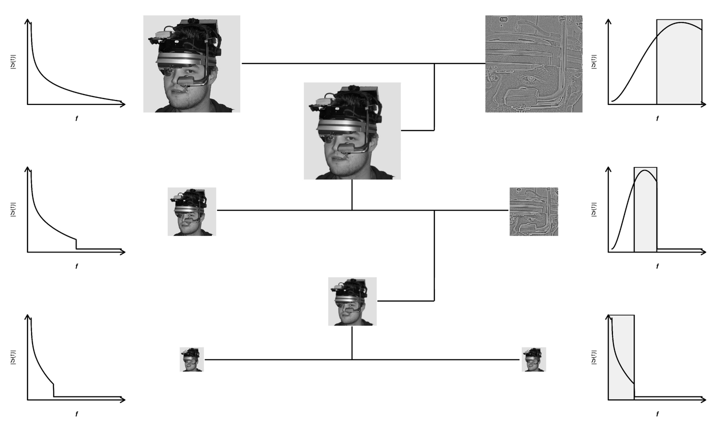


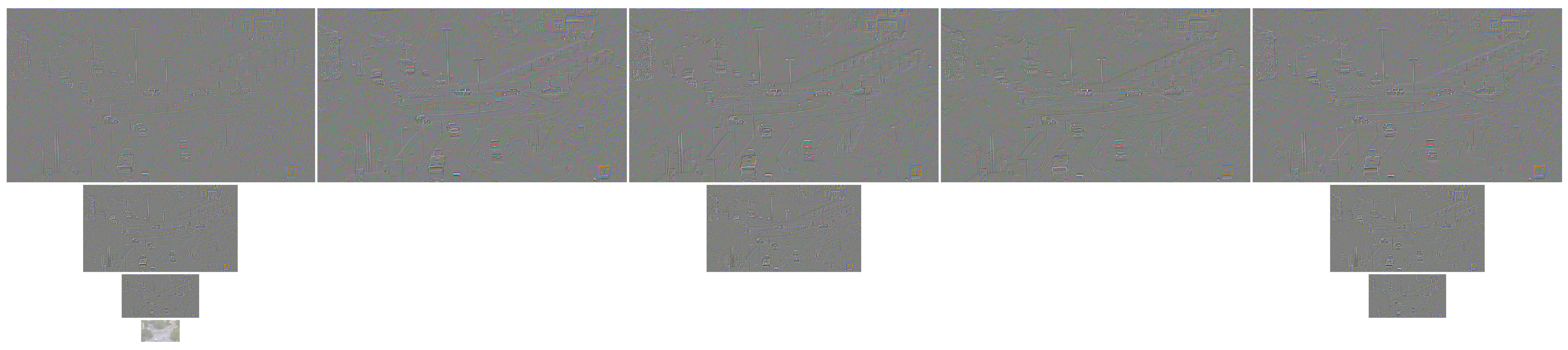
Top: Spatial foveation [1]. First, a Gaussian multi-resolution pyramid is created for each input frame (left). For each pixel in the output image (right), a cut-off frequency is specified in retinal coordinates. Pixels are either taken directly from a pyramid level (along the circles) or interpolated between two adjacent levels. Here, an exponential falloff of resolution towards the periphery is simulated, but arbitrary resolution maps can be defined.
Bottom: Temporal 'foveation' [2]. The same principle as above is extended to the temporal domain. Moving objects in the periphery are blurred or almost disappear; we could show that long-range saccades occur less often on such a gaze-contingent display [3]. Because each level has to be upsampled to full (spatio-temporal) resolution at each frame, even high-end PCs barely suffice to achieve real-time performance.



Spatio-temporal Laplacian pyramid

Foveation approaches based on Gaussian pyramids (see left) only allow to specify a cut-off frequency per pixel in retinal coordinates. We now extended our algorithm to be based on a spatio-temporal Laplacian pyramid, which allows to assign a set of (gaze-contingent) weights to individual frequency bands.
Right: A Laplacian pyramid (spatial domain only for visualisation) consists of differences of adjacent levels of a Gaussian pyramid; because levels differ only in their higher-frequency content, this yields an efficient bandpass representation.
Bottom: Example frames from a spatio-temporal Laplacian pyramid. The DC component (bottom left) represents the spatio-temporally static background.





Demo movies will be shown during the poster session!

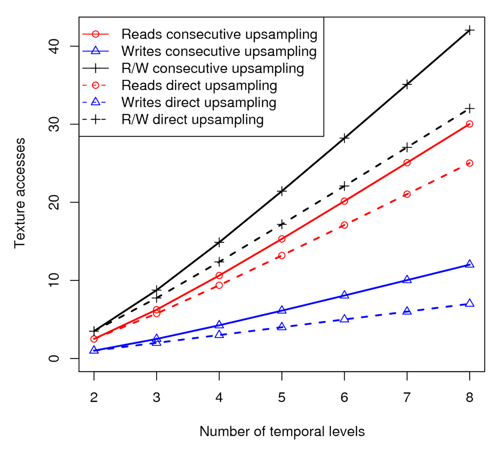
References
[1] Jeffrey S Perry and Wilson S Geisler. Gaze-contingent real-time simulation of arbitrary visual fields. In: Human Vision and Electronic Imaging vol. 4662, Bernice E Rogowitz and Thrasyvoulos N Pappas (eds.), 57-69, 2002.
[2] Martin Böhme, Michael Dorr, Thomas Martinetz, and Erhardt Barth. Gaze-contingent temporal filtering of video. Proceedings of Eye Tracking Research & Applications (ETRA), 109-115, 2006.
[3] Erhardt Barth, Michael Dorr, Martin Böhme, Karl Gegenfurtner, and Thomas Martinetz. Guiding the mind's eye: improving communication and vision by external control of the scanpath. In: Human Vision and Electronic Imaging vol 6057, Bernice E Rogowitz, Thrasyvoulos N Pappas, and Scott J Daly (eds.), 2006.

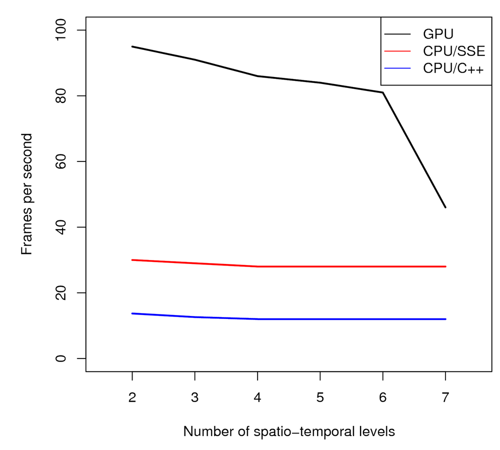
Gaze guidance

Safety-critical applications, such as car driving, could benefit from gaze guidance: the driver's attention is monitored and, if need be, directed towards a potential hazard.

We intend to use our gaze-contingent display to guide gaze by changing the saliency of a scene in real time. Based on a prediction where the subject will look (and should optimally look), distracting visual input will be suppressed (filtered), the desired location be enhanced. Bottom left: Fixation map of a dynamic scene. Bottom right: Red marker indicates gaze position; some 'hot spots' of the fixation map left (candidate locations for the next saccade) are contrast-reduced. In a preliminary experiment, we used only a CPU-implemented spatial Laplacian pyramid (with fairly long latencies) for this saliency modification; no clear guidance effect was found, but the number of saccades decreased significantly. Other modifications (temporal foveation, blinking red dots) further indicate that gaze guidance is possible in principle. With the faster GPU-based implementation and a modification of the temporal domain, we expect to achieve a continuous guidance.

Implementation





Left: Instead of iteratively upsampling temporal levels as in [2], we use 'direct upsampling' for a performance gain of about 20%.
Right: Performance measurements on a 3 GHz Intel Core 2 Duo with an NVIDIA GTX 280 graphics board. On HDTV videos (1280 x 720 pixels), the spatio-temporal Laplacian pyramid can be synthesised and reconstructed as a function of gaze at almost 100 frames per second. Hand-optimized SSE assembly routines for the CPU are about twice as fast as standard C++ code; the Graphics Processing Unit is fastest by another factor of three. Indeed, the bottleneck is the CPU converting textures; since typical videos (at 30 fps) do not need to be downsampled as often as upsampled, real-world latencies of << 10 ms can be expected. The drop in performance for the GPU on 7 levels is due to memory constraints.

Summary

- We have implemented a gaze-contingent spatio-temporal Laplacian pyramid
- Graphics Processing Unit outperforms CPU by a factor of three
- Low latency: upsampling step takes less than 1 ms
- We use this display to guide gaze by changing the saliency distribution of a scene in real time

Our research has received funding from the European Commission within the GazeCom project (IST-C-033816) of the FP6.