Fast and Novel Approach for Scale/Rotation Invariant Feature Detection

UNIVERSITÀ DEGLI STUDI DI TRENTO

Alessandro Roat, Paul Chippendale
aleroat@fbk.eu , chippendale@fbk.eu
FBK irst , Trento (Italy)

FONDAZIONE BRUNO KESSLER

TEV Technologies of Vision

Problem description:

Salient feature detection and description is an important aspect in many vision tasks: Stereo matching, object retrieval, object tracking...
The aim is to extract characteristic points of objects in a picture in a manner that they maintain coherence under scene transformation and changes of point of view.

Main solutions used:

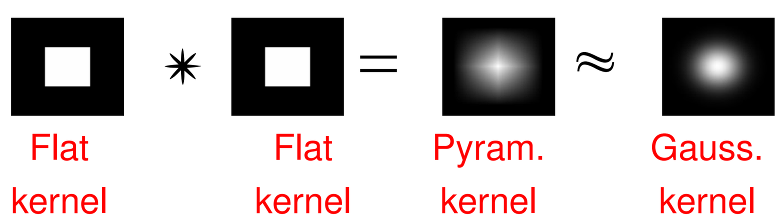
- Harris and Susan corner extractor + NCC for comparison (Fast but not resilient to rotation/scale)
- SIFT + L1 norm comparison of descriptors (Slow but resilient to rotation/scale)

Goals:

Increase the speed of the feature extraction step, using a technique similar to SIFT, whilst maintaining a good level of precision.
A multi scale approach is maintained but the detector presented here has been invented in order to be faster that SIFT using summation and multiplication.
In the encoding stage only a single *arg(x)* function is used on complex arguments in order to infer the feature orientation.
There are no multiple feature detections at different scales on the same pixel.
Scanning of the {x,y,s} volume is done more effectively without a full scale neighborhood analysis.

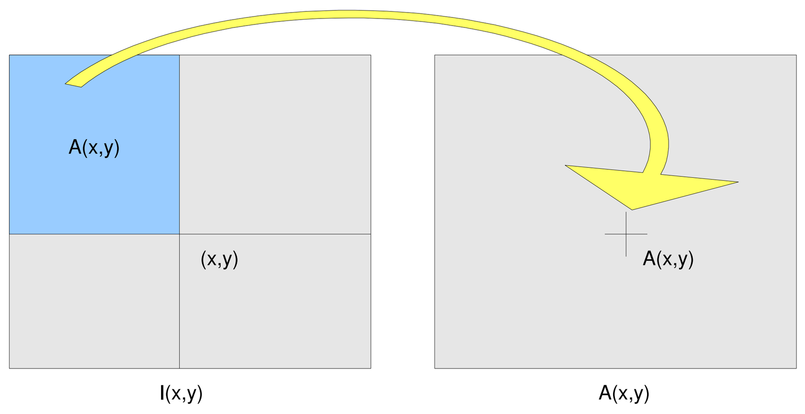
Feature detection:

Build a pyramid of smoothed images without using a Gaussian kernel.
Gaussian kernel can be approximated as a pyramid filter, that is the convolution of 2 mean windows.
Pyramidal smoothing is obtained by 2-step flat smoothing.



With further steps, a kernel more similar to Gaussian is obtained.

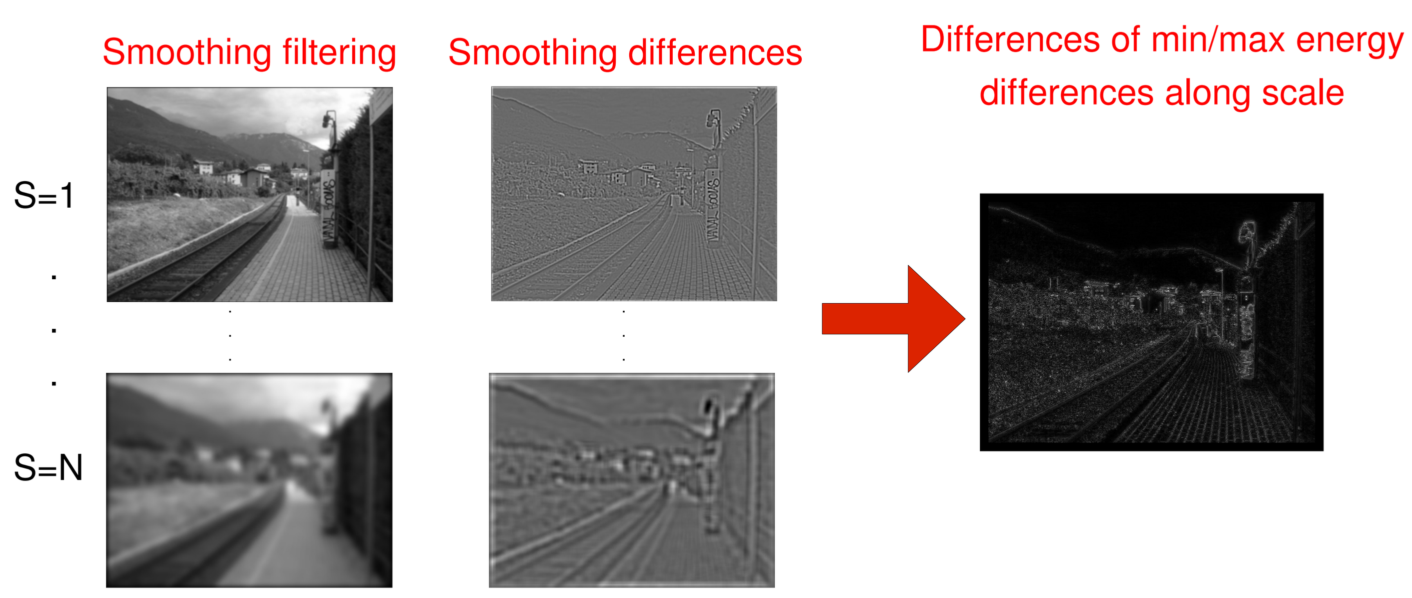
Flat smoothing is effectively implemented using integral images.



Integral of I(x,y) in a rectangular region that can be expressed as the sum/difference between 4 pixel values of A(x,y).

A pyramid of several smooth scales is built by increasing the integration window size.

For each pixel the minimum and maximum energy level is detected once on the smooth scale. Features are most likely present where the difference between the min energy and max energy is high, then a map is computed using this metric, this difference is used as a measurement of the likelihood that a pixel corresponds to a feature. Only when this difference is over a certain threshold, we check if the feature is in an local maximum, on a 4x4 window, of the likelihood function. This allows us to avoid the problem of multiple feature detection.



Changing the threshold is possible to variate the number of detected features.

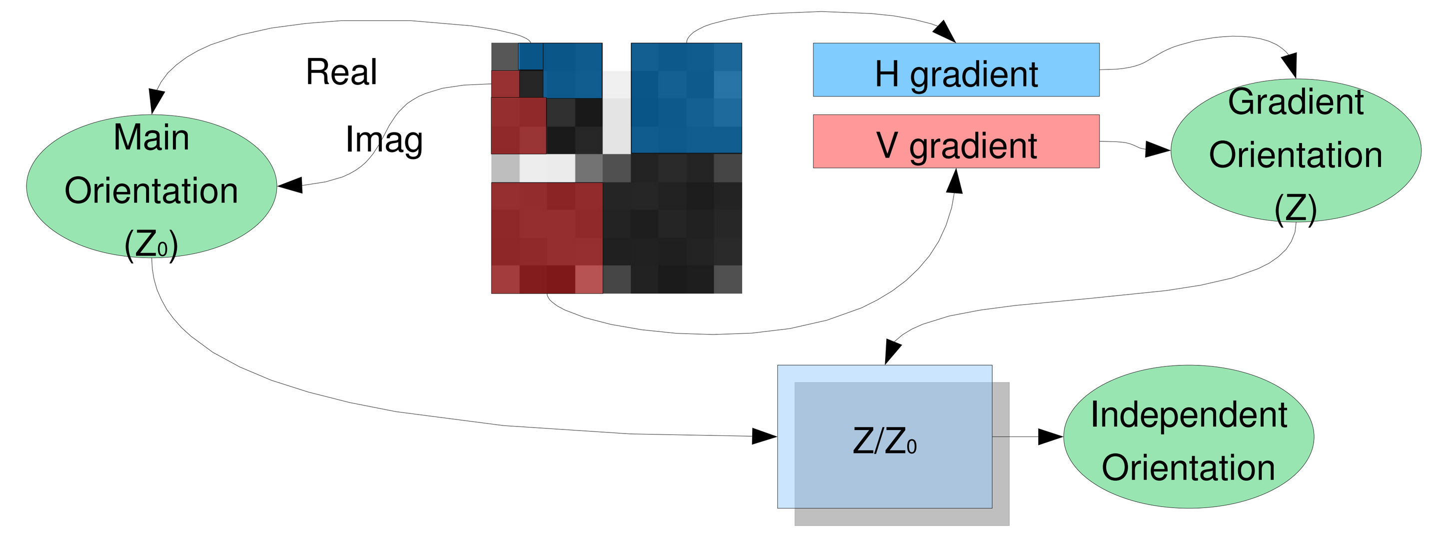
Descriptor encoding:

Features are characterized by the gradient orientation statistic of a 9x9 patch centered on the feature.
Gradient information must be transformed to obtain scale and rotation independence.

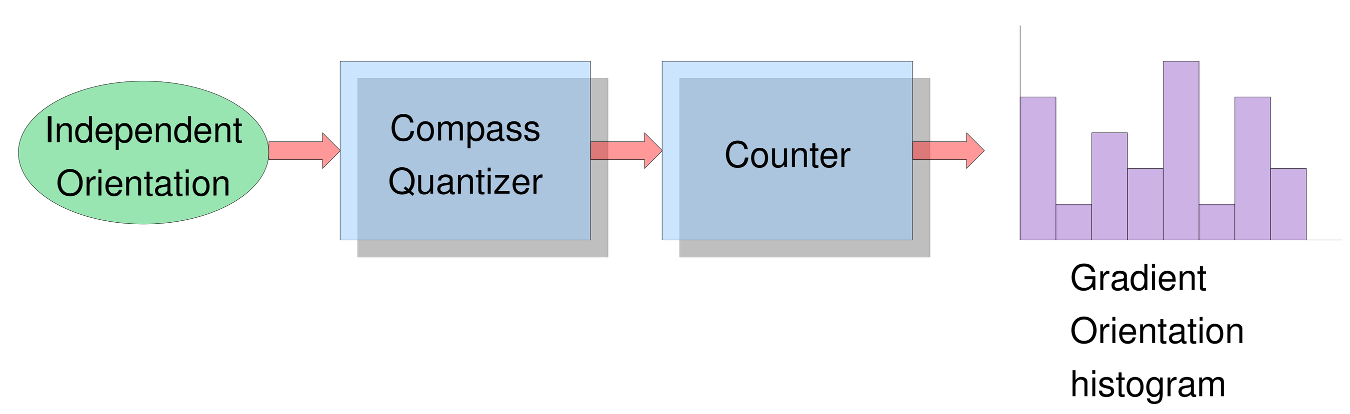
- Scale independence is obtained by sampling the patch from the highest energy level in the difference scale.
- Rotation independence is obtained through an analysis of the dominant orientation of the gradients in the patch around the feature.

A multi scale horizontal and vertical gradient are extracted using the Hadamard-Walsh transform.

Pixels (0,1) and (1,0) of transformed patch are used as Real and Imaginary parts of a complex number whose arguments represent the main orientation of the feature.
The same is done for each pixel couple in Hor./Ver. Gradient sub-patch in order to obtain gradient orientation.
Each of them is the multiplied by the conjugate of the feature orientation complex number to obtain rotation independence.

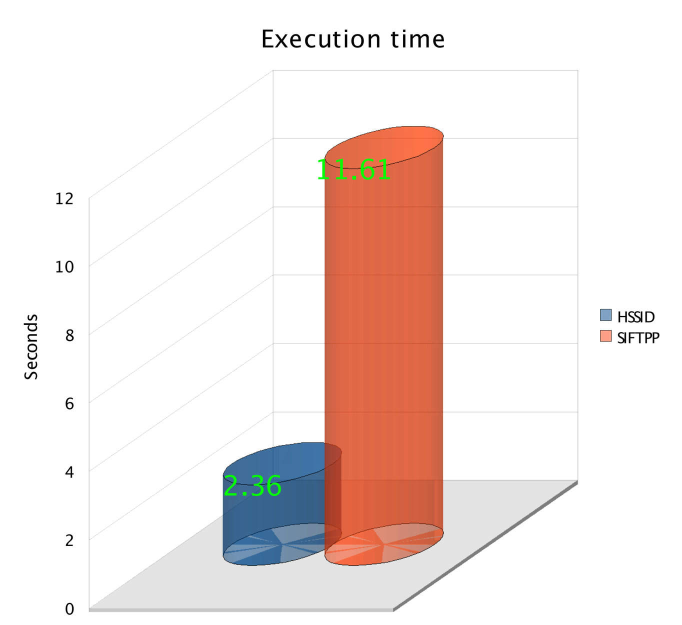


Compensated orientations are quantized in a compass-like way (8 directions) and histograms are computed at different gradient scale levels. The final descriptor is then computed by stacking the histograms into a unique vector.

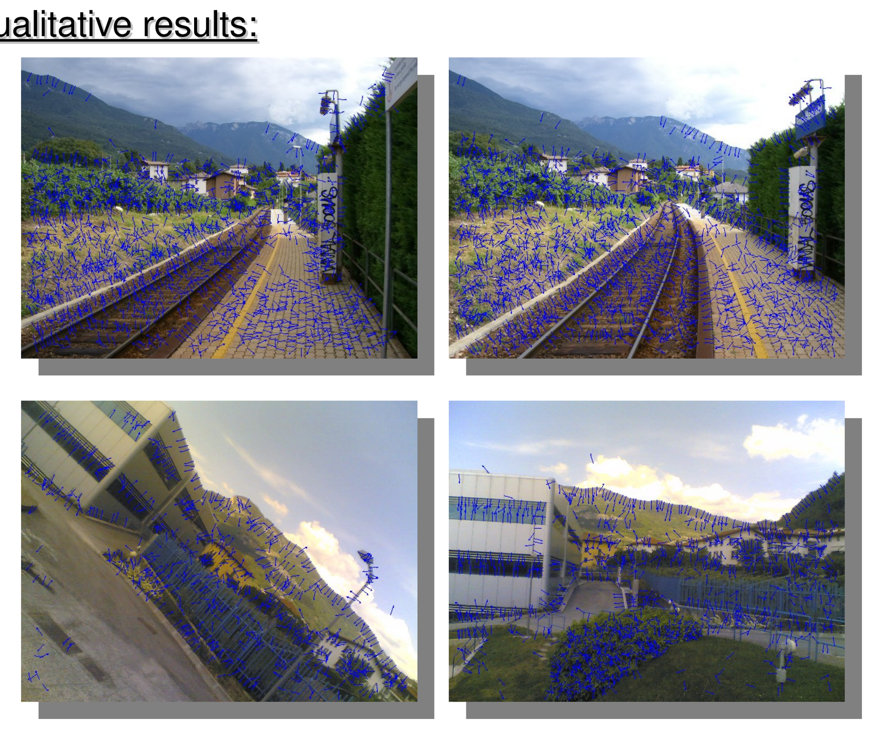


Descriptors are matched using an L1 norm.

Speed comparison:



Qualitative results:



Features are consistent on images of the same scene taken from different points of view and with slight rotation and scale variation.
Orientation of features is preserved even with large baseline displacement.
Matching performance is under evaluation ...

References :

[1] *"A Combined Corner and Edge Detector" ;* Chris Harris and Mike Stephens; Plessey Research Roke Manor, UK ; Proceedings of The Fourth Alvey Vision Conference, Manchester, pp 147-151 ; 1988

[2] "SUSAN - a new approach to low level image processing*" ; S. M. Smith and J. M. Brady*; International Journal of Computer Vision 23: 45-78 ; 1997

[3] " Distinctive image features from scale-invariant keypoints*" ; David G. Lowe*; Computer Science Department, University of British Columbia, Vancouver, B.C., Canada ; International Journal of Computer Vision, 2003