Descriptor Learning Using Convex Optimisation

MA17BTECH11002,MA17BTECH11007

March 7, 2019

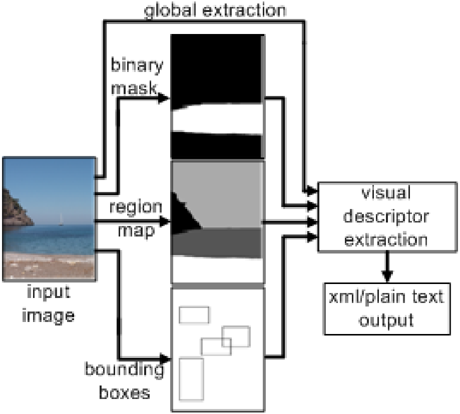
Abstract

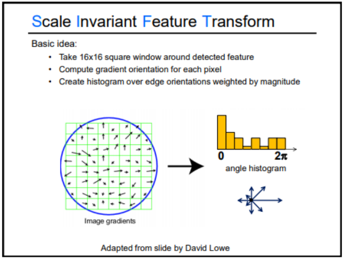
The objective of this work is to learn descriptors suitable for the sparse feature detectors used in viewpoint invariant matching.

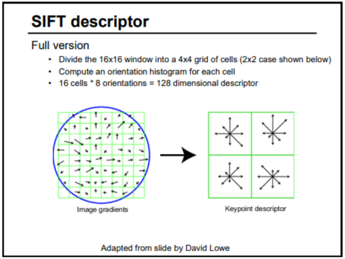
Contributions **:**

- First, it is shown that learning the pooling regions for the descriptor can be formulated as a ***Convex Optimisation*** problem selecting the regions using sparsity.
- Second, it is shown that dimensionality reduction can also be formulated as a ***ConvexOptimisation*** problem,We propose using the nuclear norm to reduce dimensionality.

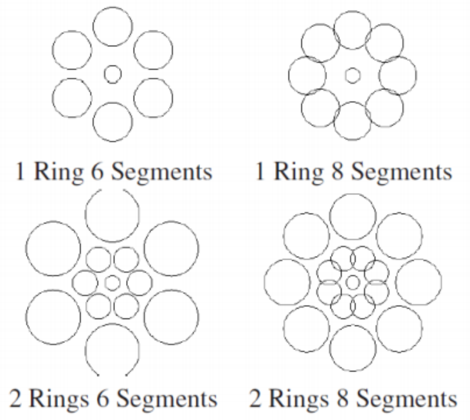
Descriptor **:**
A function that is used to describe the characteristics of an image, mainly used in stereo vision and/or computer vision.




Scale Invariant Feature Transform
Basic idea:
• Take 16x16 square window around detected feature
• Compute gradient orientation for each pixel
• Create histogram over edge orientations weighted by magnitude
0
2π
angle histogram
Image gradients
Adapted from slide by David Lowe


SIFT descriptor
Full version
• Divide the 16x16 window into a 4x4 grid of cells (2x2 case shown below)
• Compute an orientation histogram for each cell
• 16 cells * 8 orientations = 128 dimensional descriptor
Image gradients
Keypoint descriptor
Adapted from slide by David Lowe

Circular gradient binning



Daisy

Computation

- Gaussian smoothing is applied to patch image.
- Intensity gradient is calculated at every pixel(8 possible directions, hence p=8).
- Normalise the gradient with T(x).
- Project it wrt a matrix to reduce dimensionality.

Computation

$$\boldsymbol{T(x) = (mean(g(x)) + std(g(x)))/p}$$

- g : magnitude of gradient after normalisation.
- v : determines the amount of cropping.

$$\boldsymbol{\psi_i(x) = min\left\{\tilde{\psi}_i(x)/T(x), 1\right\}} \forall i.$$

let ψ be the descriptor defined by PRs pool subset encoded by the weight vector w_i:

$$\boldsymbol{\psi_{i,j,c}(x) = \sqrt{w_i}\phi_{i,j,c}(x)} \quad ; w_i \geq 0$$

$$\boldsymbol{d(x,y) + 1 < d(u,v)} \quad \forall (x,y) \in P, (u,v) \in N$$

where P and N are the training sets of positive and negative feature pairs

Computation

$d(x, y)$ is the distance between x and y...
d(x, y) $=||\psi(x) - \psi(y)||^2$
Some manipulation on the equations and we get the convex optimisation problem :
where d_w is the squaredL^2 distance in the projected space:

$d_w(x, y) =$ || W $\psi(x)$ - W $\psi(y)||^2$

$=$ ($\psi(x)$ - $\psi(y))^T W^T$ W ($\psi(x)$ - $\psi(y)$)

$= \theta(x, y)^T A\theta(x, y),$

with $\theta(x, y) = \psi(x) \;\; \psi(y)$, and A $=W^T$ W is the Mahalanobis matrix

Formulation

Hence the Convex Optimisation problem we are left with simply is :
$\boldsymbol{argmin_{\eta,b}}\ \boldsymbol{\Sigma_x}\,\mathrm{b(x)}\ \max\left\{\boldsymbol{min_{y\in P(x)} d_\eta(x,y) - min_{u\in N(x)} d_\eta(x,u) + 1, 0}\right\}$

Subjected to the constraint(s) :

$$d(x,y) + 1 < d(u,v)$$

Proof of Convexity

Here the Constraint has the terms like $d(\eta)(x,y)$ which is the distance (2 norm)..
It is the sum of squares of some variables. We know that squares are convex and sum of convex function is convex. Adding 1 doesn't change the convexity.

Hence, $d(\eta)(x,y)$ is convex.
Then we have maximum of such functions which is also convex.

So our objective and constraint functions are convex.

RDA

To handle such very large training sets, we propose to use ***RegularisedDualAveraging*(*RDA*)**, RDA is a stochastic proximal gradient method effective for problems of the form:

$$min_w \frac{1}{T} \sum_{t=1}^{T} f(w, z_t) + R(w)$$

Compared to other proximimty methods, RDA uses more tighter threshold.

Results

We compare our learnt descriptors with those of in two scenarios: (i) learning pooling regions and (ii) learning discriminative dimensionality reduction on top of learnt PRs. In both cases the proposed framework significantly outperforms the state of the art, reducing the error rate by up to 40%

The End