Descriptor Learning Using Convex Optimisation

MA17BTECH11002,MA17BTECH11007

March 4, 2019

Abstract

The objective of this work is to learn descriptors suitable for the sparse feature detectors used in viewpoint invariant matching. We make a number of novel contributions towards this goal: 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 convex optimisation problem,We propose using the nuclear norm to reduce dimensionality.

Intro

Descriptor **:**
A function that is used to describe the charachteristics of an image, mainly used in stereo vision and/or computer vision. Basically there have been many algorithms to perform this, with most of them being based on ***SIFT***. But the problem with them is, they have non convex forms, often resulting in non optimal solutions.
Here we try to address that effeiently by formulating it as a Convex Optimisation Problem.
Here we show that normalisation need not be done after descriptors are computed, instead we can get the normalisation factor directly from the image patch

Related works:

Some Conventional feature descriptors...
SIFT : Uses rectangular regions.
DAISY: a set of multi-size circular regions grouped into rings.

Some Approaches to reduce dimensionality...
PCA : Prioritises features which for which the descriptor values differ highly for two elements of different sets.

LDA: This works by prioritising features which maximise the difference between two images of different sets while making sure that those of same set are as close as possible.

It is shown in some previous works that PCA outperforms LCA if Pooling Regions are optimised.To encourage dimensionality reduction, we utilise the matrix nuclear norm.But this results in Convex but non smooth objective.

Computation

Gaussian smoothing is applied to patch image. Intensity gradient is calculated at every pixel(8 possible directions, hence p=8).
Pooling is applied separately to each feature channel, which results in the descriptor vector $\psi_1(x)$ with dimensionality pq, where q is the number of PRs.
The output of each filter is divided by the pre-computed normalisation factor T(x) and thresholded to obtain responses $\tilde{\psi}$(x) invariant to intensity changes and robust to outliers.Then we project $\tilde{\psi}(x)$ through a matrix which results in redcuction of dimension to get $\psi(x)$.

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

where g is the gradient magnitude is the normalisation factor with v determining the amount of cropping.

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

let ψ be the descriptor defined by PRs pool subset encoded by the weight vector w_i:
$\psi_{i,j,c}(x) = \sqrt{w_i}\phi_{i,j,c}(x) \quad ; w_i \geq 0$

$d(x,y) + 1 < d(u,v)\forall(x,y) \in P, (u,v) \in N$
where P and N are the training sets of positive and negative feature pairs

$d(x, y)$ 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

To handle such very large training sets, we propose to use Regularised Dual Averaging (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%

Consider image pairs randomly sampled from the dataset, for which the homographies are automatically estimated.

For each feature x of one image, we compute the sets P(x) and N(x) of putative positive and negative matches in another image based on the homographies and the region overlap criterion . We aim at learning a descriptor such that the NN of x is a positive match from P(x). To account for the cases where x can not be matched based on its appearance, we introduce a binary latent variable b(x) which equals 0 iff the match can not be established. This leads to the optimisation problem:

arg $min_{\eta,b}$ Σ_xb(x) max $\left\{min_{y\in P(x)}d_\eta(x,y) - min_{u\in N(x)}d_\eta(x,u)+1,0\right\}+$ $R(\eta)$.

s.t. $b(x)\in 0,1$, $\Sigma_x b(x) = K$

The End