Vector Quantization of Stars and Galaxy for Dark Matter Mapping Applications

Andriyan B. Suksmono

STEI-ITB

Objective

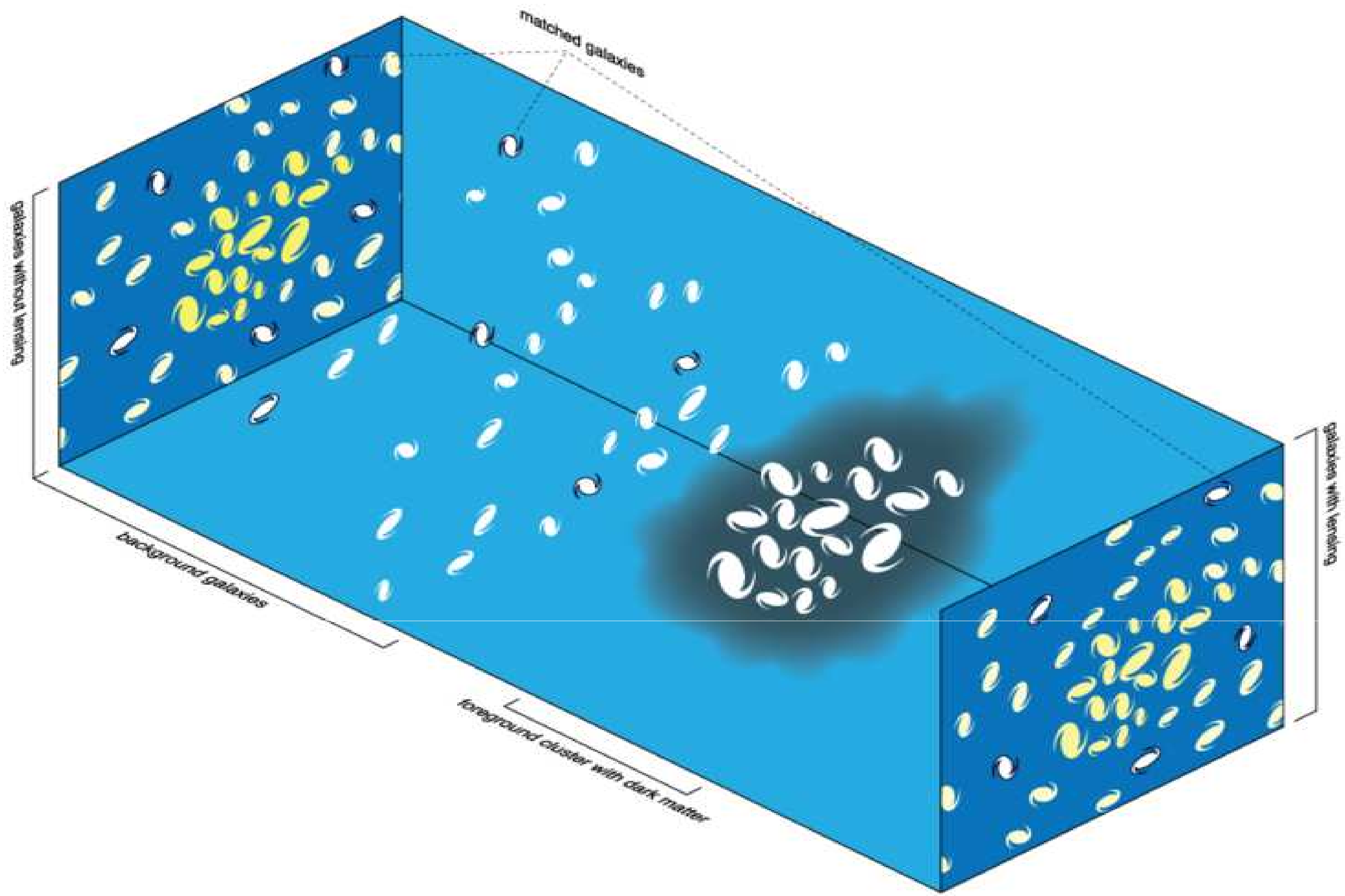
- Develop A New Method to Measure A Galaxy Ellipticity For Weak Lensing Measurement
  - Ellipticity Distribution can be Used to Infer and Map of Unseen Distributed Matters
  - Need for Accurate Measurement

Proposed Method

- Vector Quantization
  - Basically a table-lookup method:
    - directionally stacked images, split into a number of reference vectors
    - Use the reference to measure ellipticity by best-matching
  - Two possible VQ techniques are investigated
    - Direct VQ of Raw Images
    - VQ on Image Parameters (FFT Coefficients)

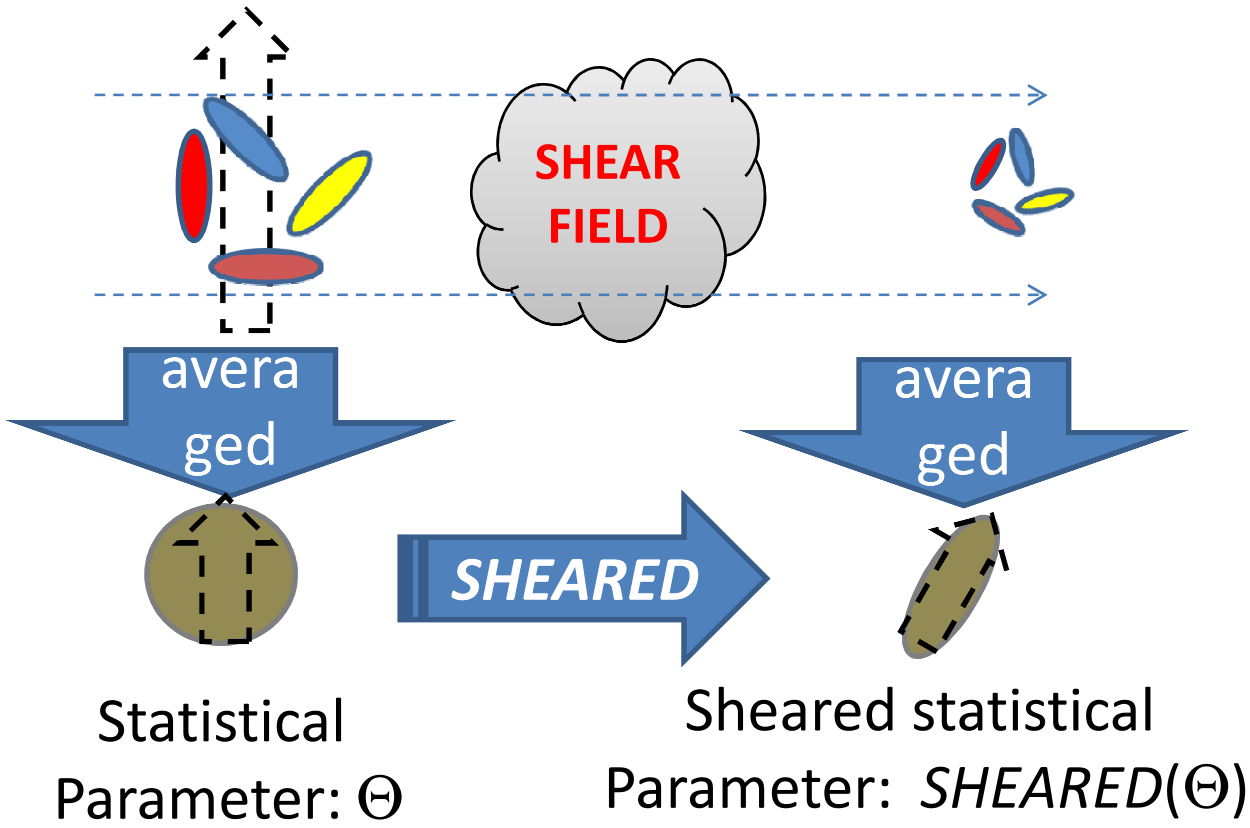
Background

- The matters that we see in daily life; all object around us, moon, planets, stars, and galaxy, are only a small parts (~5%) of the universe. According to the most recent Astrophysics/ Cosmulogy findings, most of the universe are consisting of dark energy (~75%) and (cold-) darkmatter (20%).
- Although darkmatter cannot be observed directly, its pressence causing space-time curvature, can be detected by analysing the changes of its neighbouring objects.
- Accurate measurement of galaxy shape, i.e. the ellipticity and related parameters, caused by weak gravitational lensing is a powerful method to map the distribution of the darkmatter.



- *Figure 1.Gravitational Field of groups of galaxy changes the shape of background galaxy. Darkmatter, although cannot be seen, change space-time curvature around them in a similar way, so that its existence and distribution can be map by measuring the distribution of the ellipticity [Copyright: Wikipaedia.org].*

How to Measure ?



- *Figure 2. Random orientation of galaxy yields zero ellipticity value when the space-time not affected by mass. The presence of darkmatter induces shear field, so that averaging ellipticity values in a region gives a small but non-zero residual shear value.*

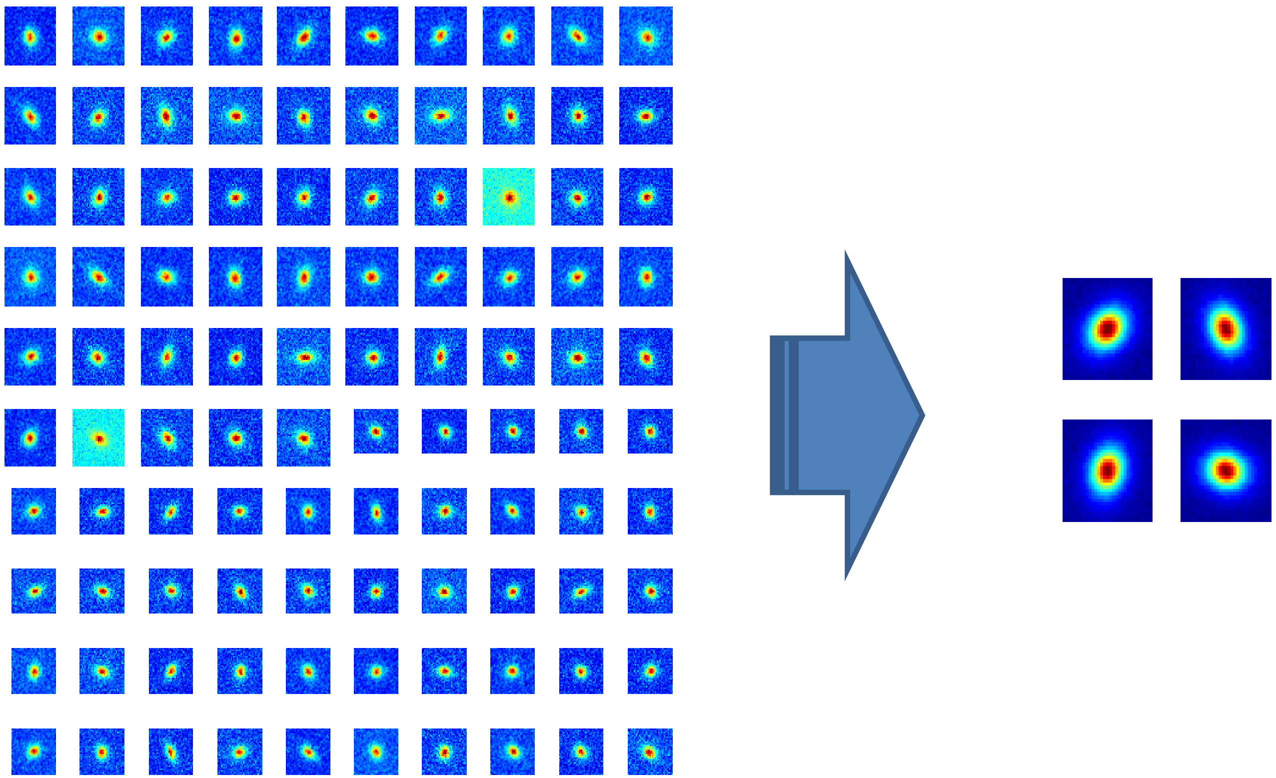
The Challenge

- Small changes in ellipticity needs accurate measurement of the galaxy ellipticity.
- Non-ideal condition in the observation:
  1. blurring/smearing caused by non-ideal optical component and atmospheric disturbance,
  2. pixelation effect by limited capacbility of the sensor/CCD, and
  3. unavoidable thermal noise of the instruments.

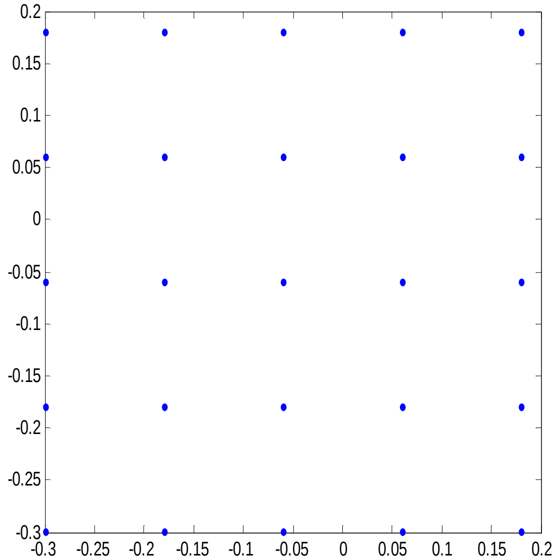
Solution

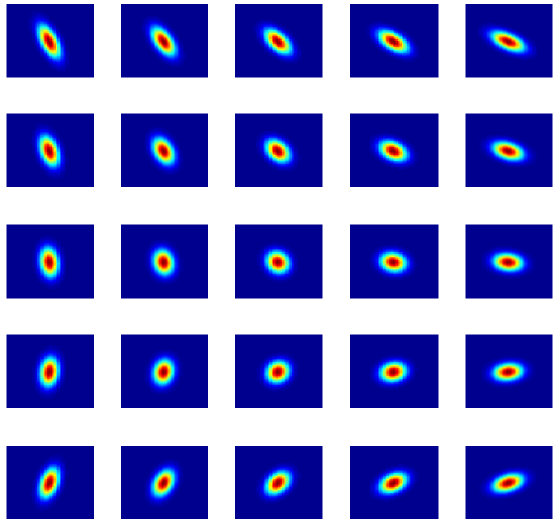
- We propose VQ (Vector Quantization) to measure the ellipticity is based on the following considerations
  - Codebook construction is performed by clusterring and stacking. Stacking will reduce the variance or noise energy, proportional to the number of objects in the cluster.
  - The accuracy is scallable,i.e, the larger the codebook size, the smaller the difference (error) between the actual value and the prototype.
  - It is possible to lower the noise floor by adding more member in a cluster.

VQ of Observed Images



Synthesize the Codebook





The 2D ellipticity space is partitioned into K-subsapce, then the center of each partition will be use to construct the prototype.

Experiments

Case-1: Noiseless-VQ

- Two basic VQ are performed:
  1. VQ on the values of ellipticity (VQE)
  2. VQ on generated images (VQI)
- It is expected that, when the codebook size is increased, then
  – VQE: obviously (based on Rate-Distortion Theory), resolution will increase MSE will decrease
  – VQI: resolution will increased/ MSE will decrease -> to be confirmed, since now the ellipticity values has been mapped into elliptic-gaussian function.
  – The difference of the two will be observe: VQE ~ VQI, VQE > VQI, or VQE < VQI?

Case-2: Noisy VQ

- VQ, particularly VQI, for data with noise will be evaluated:
  - Effect of noise power on MSE
  - Identify the “filtering” effect, in what noise regime it is effective:
    - Low ?
    - Medium ?
    - High ?

Results: noiseless data

No.	CD_SIZE	CR	N_TRAIN	MSE	
				ELLIPTICITY	IMAGE
1	8	32	128	0.042070	0.042457
2	16	16		0.027120	0.025043
3	32	8		0.014854	0.015216
4	64	4		0.007283	0.008895
5	8	64	512	0.042888	0.045471
6	16	32		0.030825	0.030768
7	32	16		0.020123	0.020767
8	64	8		0.014602	0.015447
9	128	4		0.008976	0.009585
10	256	2		0.004728	0.006386

- Two sets of VQ with N-train 128 and 512 are conducted.
- CR: compression ratio, ratio of N_TRAIN to Codebook Size

Analysis

- The table shows:
  - VQI and VQE are comparable
  - MSE reduced as codebook size increased, or compression ratio decreased.
  - For the two sets with different N-train, the MSE value on the same CR are comparable

VQ on Noisy image

No.	Noise	MSE		
		QM	VQ	VQ_QMFFT
1	10%	0.007719	0.011242	0.009146
2	20%	0.014278	0.015202	0.010331
3	30%	0.020728	0.022004	0.010974
4	40%	0.031968	0.029875	0.010350
5	50%	0.034030	0.032161	0.013327
6	60%	0.044874	0.038375	0.012630
7	70%	0.057340	0.052018	0.016543
8	80%	0.060293	0.055822	0.015657
9	90%	0.071178	0.065540	0.019778
10	100%	0.085677	0.071020	0.021548

- NTRAIN=128, CDSIZE=64
- Compare VQ with existing QM (Quadrature Moments) Method

Analysis

- Simulation Results Indicates:
  - Upto 30% noise energy, QM perform better than VQ.
  - In high noise regime (>30%), VQ perform better: directional stacking start to works removing the noise.
  - In practice (benchmark data), low MSE is expected (<0.02). Direct VQ possibly becomes impractical.
- Why it doesn't work well?
  - Ellipticity values are determined by QM on noisy image.
- Possible Improvements:
  - Ellipticity or related parameters should has been measured based on clean images: Use synthetic codebook
  - On clusterring:
    - Better to use QM-params, instead of ellipticity (linearity issues)
    - Better to use feature that not-sensitive to centroid: Absolute FFT of the image

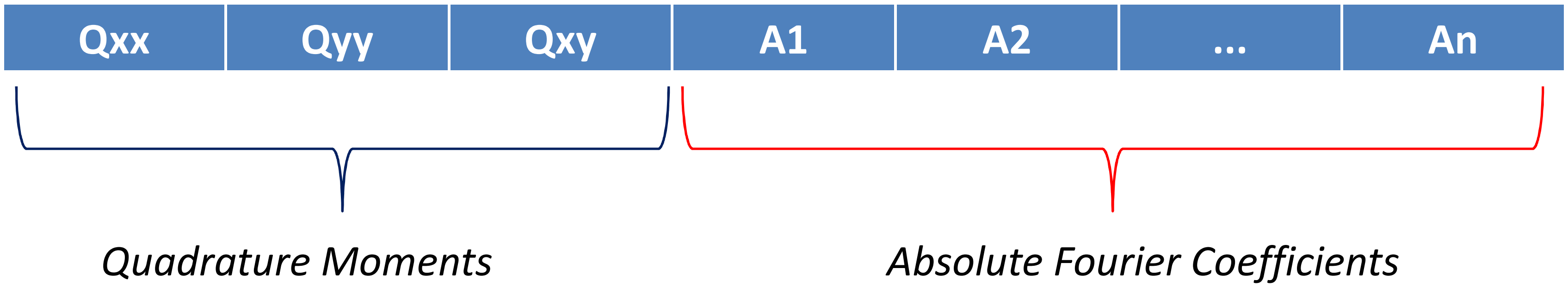
Further Improvements

FFT Features

Scenarios

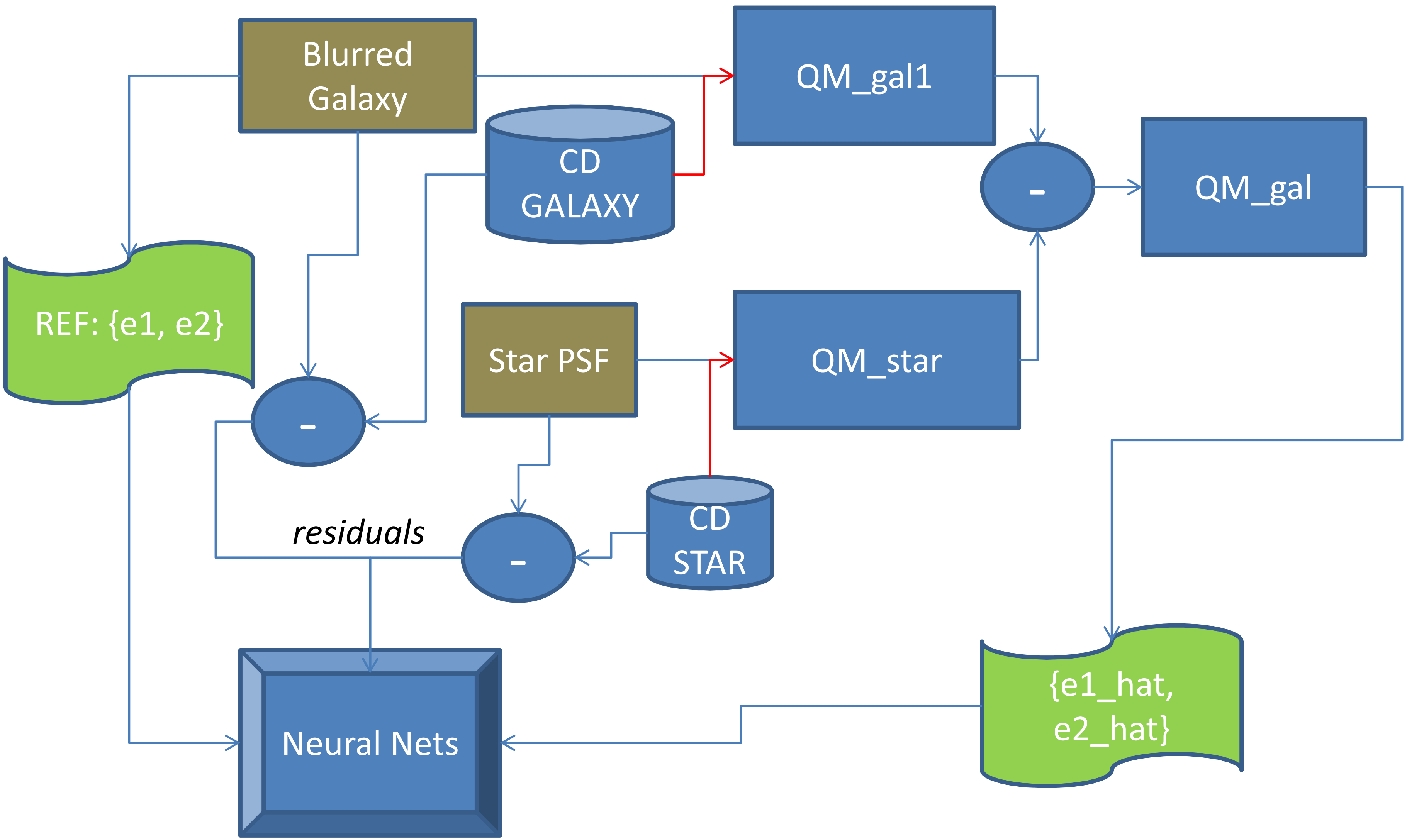
- Embedding elliptic parameters on the image data/feature:
  - Reason:
    - direct measurement on codebook entry is not accurate for high-noise regime
    - Better to generate “synthetic” codebook where the ellipticity is known beforehand
  - Absolute FFT feature: reduce the image into a few parameters, non-sensitive to centroid

Feature



- Feature consisting of two parts
  - Absolute Fourier Coefficients
    - Only half is required, due to symmetricity and no-centroiding-problem aspect; representing elliptical geometry or shape of the image
    - Used as a “key” to retrieve codebook entry
  - QM is embedded in the feature
    - Better representing ellipticity

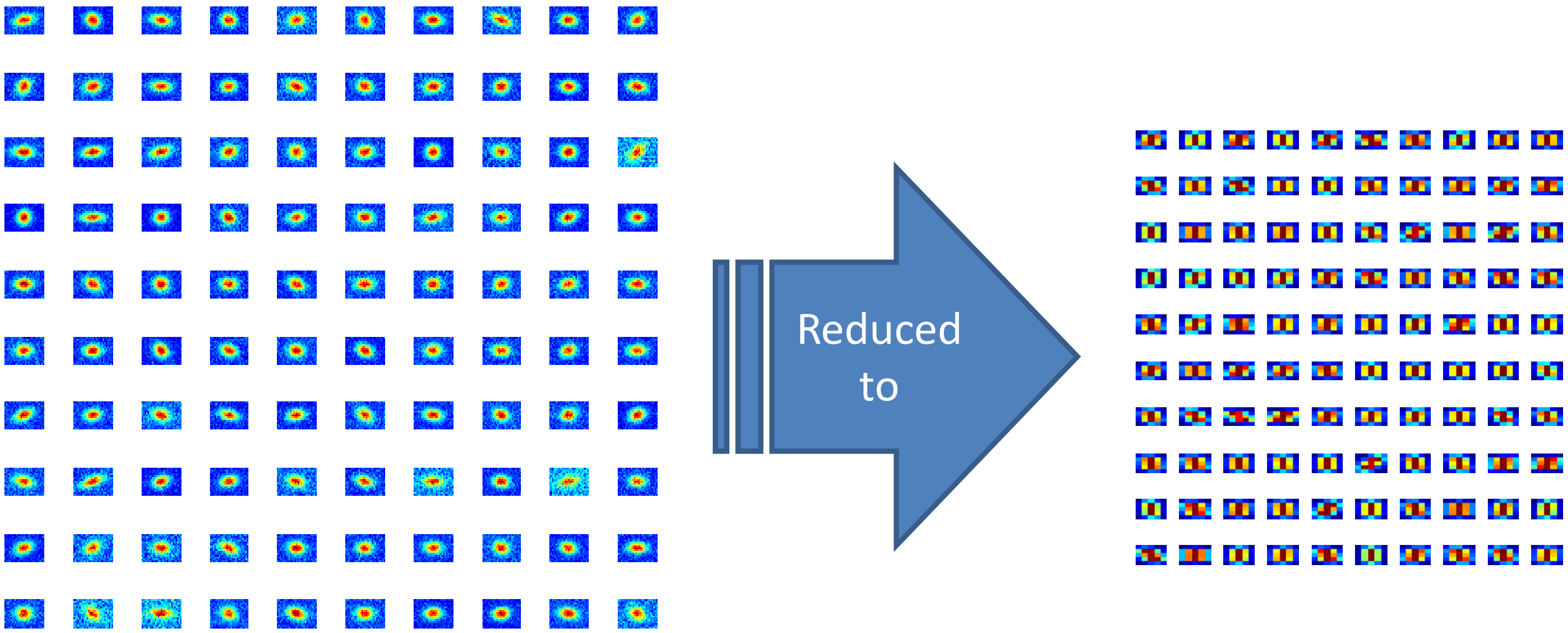
VQ_QMFFT


Blurred Galaxy
QM_gal1
CD GALAXY
-
QM_gal
REF: {e1, e2}
Star PSF
QM_star
-
CD STAR
residuals
-
Neural Nets
{e1_hat, e2_hat}

Simulation Steps

- Generate random numbers: E:={e_g, e_s}, E1={e_g1, e_s1}
- VECTOR QUANTIZATION STAGE:
  - VQ of the Galaxy:
    - Generate gaussian galaxy from E
    - Calculate abs(FFT) of the galaxy: AGALS
    - Calculte QM of the galaxy
    - Construct galaxy feature training set: XGAL{QM_gals, 100*AGALS}
    - Construct Galaxy Codebook: GAL_ctrs
  - VQ of the STARS
    - Generate Moffat -stars from E
    - Calculate abs(FFT) of the stars: ASTARS
    - Calculate QM of the stars
    - Construct star feature straining set: XSTAR{QM_stars 100*ASTARS}
    - Construct Star Codebook: STAR_ctrs
- EVALUATION STAGE
  - Generate sersic -galaxy and moffat-stars from E1: fgal, fstar
  - Simulate degradation: bgal=fgal*fstar+noise
  - Normalization of the object (bgal)
  - Calculate feature: fft of the bgal AGALSC
  - Use features to retrieve codebook-entries: get VQ of QM params: QM_gal
  - Do similar things with the star: QM_gal
  - Use QM_gal to correct QM_gal, calculate ellipticitieas

Codebook of FFT Coefficients (Abs)



Use (half) magnitude coefficients of FFT2

3x5 = 15 length feature (instead of 50x50=2500)

Preliminary Results

- Simulation Parameters:
  - 400 galaxy
  - 100 size codebook
  - Noise Variance: 0.01
- Theoretical RMS: 0.0100
- RMS of DirectVQ: 0.0263
- RMS of VQ_QMFFT: **0.0142**

Summary

- VQ of the Image Parameters Outperform Direct VQ
  - More Stable Ellipticity Measure in Frequency Domain
  - Selecting Fewer Dominant Parameter
    - implies Low Pass Filtering -> Reduce Noise
    - Increase Computational Speed

Next Steps

- Evaluate both of Direct VQ and FFT-VQ on Benchmark Data
- Fine Tuning the Performance using Neural Networks
- Write A Comprehensive Report
  - Submit to A Journal

Thank You