Q1)The joint distribution of Y and Z can be found using Bayes' rule:

$$p(y,z) = p(y|z) * p(z)$$

Substituting the given distributions:

$$\mathrm{p(Y,Z) = N(WZ\ +\ b, \Sigma Y|Z)\ *\ N(\mu z, \Sigma z)}$$

Expanding the product and completing the square:

$$\mathrm{p(Y,Z) = (2\pi)^{-\frac{k}{2}} * |\Sigma Y|Z|^{-\frac{1}{2}} *\ exp[-1/2 * (Y - WZ - b)' * (\Sigma Y|Z)^\wedge - 1 * (Y - WZ - b)]\ \ *\ (2\pi)^{-\frac{m}{2}}\ *\ |\Sigma z|^{-\frac{1}{2}}\ *\ exp[-1/2 * (Z - \mu z)' * (\Sigma z)^\wedge - 1 * (Z - \mu z)]}$$

where k is the dimension of Y, m is the dimension of Z.

Taking logarithm and simplifying:

$$\mathrm{\log p(Y,Z) = -1/2 * \log|\Sigma Y|Z|\ -\ 1/2 * \log|\Sigma z| -\ 1/2 * (Y - WZ - b)' * (\Sigma Y|Z)^{-1} * (Y - WZ - b) -\ 1/2 * (Z - \mu z)' * (\Sigma z)^{-1} * (Z - \mu z) -\ k/2 * \log(2\pi)\ -\ m/2 * \log(2\pi)}$$

Comparing with the standard form of a multivariate normal distribution:

$$\mathrm{\log p(Y,Z) = -1/2 * [\ [Y;\ Z] -\ [W';\ I][W\ ,\Sigma WZ\ ;\ \Sigma WZ', \Sigma z]^{-1}[W\ ;\ \Sigma WZ\ ;\ \Sigma WZ', I][Y;\ Z] +\ const}$$

Therefore, the joint distribution of Y and Z is also a multivariate normal distribution:

$$p(Y, Z) = N([W';\ I][W, \Sigma WZ;\ \Sigma WZ', \Sigma z]^{-1}\ [W;\ \Sigma WZ;\ \Sigma WZ', I]^{-1}[b;\ \mu z], [\Sigma Y|Z, \Sigma Y|Z * W';\ W * \Sigma Y|Z, \Sigma Y|Z\ +\ \Sigma WZ])$$

$[W';\ I]$ is the matrix that obtained by stacking W vertically on top of the identity matrix, and similarly for$[W, \Sigma WZ;\ \Sigma WZ', \Sigma z]$.

The mean vector is given by$[W';\ I][W, \Sigma WZ;\ \Sigma WZ', \Sigma z]^{-1}\ [W;\ \Sigma WZ;\ \Sigma WZ', I]^{-1}[b;\ \mu z]$, and the covariance matrix is given by the block matrix $[\Sigma Y|Z, \Sigma Y|Z * W';\ W * \Sigma Y|Z, \Sigma Y|Z\ +\ \Sigma WZ]$.

Therefore, the parameters of the joint distribution are:

$$\mu Y, Z = [W';\ I][W, \Sigma WZ;\ \Sigma WZ', \Sigma z]^{-1}\ [W;\ \Sigma WZ;\ \Sigma WZ', I]^{-1}[b;\ \mu z]$$

$$\Sigma Y, Z = [\Sigma Y|Z, \Sigma Y|Z * W';\ W * \Sigma Y|Z, \Sigma Y|Z\ +\ \Sigma WZ]$$

The posterior distribution of Z given Y can be found using Bayes' theorem:

$$p(Z|Y) = p(Y|Z) * p(Z)/p(Y)$$

where p(Y) is a normalization constant. We can recognize p(Y) as the marginal likelihood obtained by integrating out Z from the joint distribution:

$$p(Y) = \int\ p(Y, Z)dZ$$

Using properties of multivariate normal distributions, we can show that both p(Z|Y) and its parameters are also normal:

$$\mu Z|Y = \mu z\ +\ \Sigma zW'(\Sigma Y|Z\ +\ W\Sigma zW')^{-1}(y\ -\ W\mu z\ -\ b)$$

$$\Sigma Z|Y = \Sigma z\ -\ \Sigma zW'(\Sigma Y|Z\ +\ W\Sigma zW')^{-1}W\Sigma z$$

Therefore, the posterior distribution of Z given Y is also normal with mean vector $\mu Z|Y$ and covariance matrix $\Sigma Z|Y$.

Q2) The posterior distribution is not necessarily normal, but it can be represented as a GMM. The parameters of this GMM depend on the parameters of the prior distribution and the likelihood function.

If the prior distribution is a GMM with K components, each with mean μk and covariance matrix Σk, and weights wk, and $pY|Z$ is a normal distribution with mean μy and covariance matrix Σy, then the posterior distribution can be represented as a GMM with K components, each with mean $\mu k'$ and covariance matrix $\Sigma k'$, and weights wk', where:

$$\mu k' = (\Sigma k^{-1} + \Sigma y^{-1})^{-1}(\Sigma k^{-1}\mu k + \Sigma y^{-1}\,\mu y)$$

$$\Sigma k' = (\Sigma k^{-1} + \Sigma y^{-1})^{-1}$$

$$wk' = wk\frac{p(y|zk)}{\sum j\ wj\ p(y|zj)}$$

where $p(y|zk)$ is the probability density function of $pY|Z$ evaluated at y for component k.

Therefore, if we know the parameters of the prior distribution and $pY|Z$, we can compute the parameters of the posterior GMM using these formulas.

Q3)

GMM (Gaussian Mixture Model) is a preferred prior distribution for Z (the distribution of patches of an image) because it can model complex distributions with multiple modes and capture the variability in the data. GMM assumes that the data is generated from a mixture of Gaussian distributions, each with its own mean and variance. This allows for more flexibility in modeling the data compared to a single Gaussian distribution.

Furthermore, GMM can be easily trained using an expectation-maximization algorithm, which iteratively estimates the parameters of the model. This makes it computationally efficient and scalable for large datasets.

In image processing, GMM has been widely used for tasks such as image segmentation, object recognition, and texture analysis. By modeling the distribution of patches in an image using GMM, we can identify regions with similar characteristics and extract meaningful features for further analysis.

Overall, GMM is a powerful tool for modeling complex distributions and has proven to be effective in various applications in image processing.