1) Suppose two of the eigenspace curves built by the routine called EIGENSPACE_LEARN in the text book intersect in some points. What does this mean in terms of identification? Answer: An eigenspace curve represents a number of different views of a single object. If the two curves intersect it means that there are a number of views for which the two objects look the same and can not be distinguished, that is the images for these views have approximately the same pixels. ½ mark
2) For two variables what does it mean if the covariance of (x,y) is a large positive number, and what does it means if the covariance of (x,y) is zero. Answer: If the co-variance is a large positive number the two variables are strongly correlated in the positive sense; that is if one variable has a given value the other variable is likely to have close to the same value. If the co-variance is zero then the two variables are uncorrelated, which means that the value of one of the variables gives us no information or no prediction about the value of the other variable. 1 mark
3) There is an N by N image which is compared to a database of images of the same size by a) sum of squared differences b) correlation c) eigenspace method. What is the complexity of each approach? Why would one method be used over the other? Answer: Assume each image has N by N pixels, and there are m images in the database. For methods a) and b) the complexity is $O(m N^2)$. For the eigenspace approach the complexity is O(m k) where k is the number of eigenvectors that are used in the new basis. Normally $k \ll N^2$ so that is why the eigenspace approach is superior to sum of squared differences and correlation. 1 ½ mark
4) For which types of images will the eigenspace approach work well, and for which type will the eigenspace approach not work well? Answer: If the images have a lot of redundancy, that is parts of all the images are very similar, such as faces in a standard position, then the eigenspace approach will work well. If the images have no redundancy, for example, each image is a set of random pixels then the eigenspace approach will not work well. 1 mark
5) Assume you are estimating the camera parameters using the projection matrix approach. The first thing you need to do is create a 3D calibration pattern, and take a number of images of that pattern with the camera that is to be calibrated. This is followed by two other steps which are necessary to actually find the camera calibration. What are these two steps? Answer: Assuming we know the 2D pixel

co-ordinates of the projections of the known 3D points on the calibration pattern. First you compute the projection matrix M that transforms these 3D points to 2D pixels, and then you use the characteristics of the rotation matrix to find the intrinsic and extrinsic parameters that are in M, that is to find Mint and Mext, where it is the case that Mint Mext = M. 1 mark

6) In one or two sentences define the correspondence problem and the reconstruction problem. Answer: Assume there are two images. The correspondence problem is to decide which image pixels correspond to the same 3D point in space. The reconstruction problem is to take these correspondences along with whatever information is known about the intrinsic and extrinsic camera parameters and to compute the actual 3D co-ordinates of that point in space. 1 mark
7) In a simple stereo system there are two cameras. Assume both cameras have the same focal length f; that the baseline of the system is T, and that for a given point with a depth value of Z, the disparity is d. Write down the equation of the depth Z as a function of d, T, and f. What is one advantage of having a large baseline b, and what is one disadvantage of a large baseline b. Answer: The simple stereo equation is $Z = f\,T / d$. The advantage of large baseline is more accurate estimation of the depth, and the disadvantage is that the correspondence problem is more difficult. 1 mark
8) There are generally two classes of correspondence algorithms. Name the two classes. One class usually produces more 3D data points than the other class. Which of the two classes produces more 3D data points, and which produces fewer 3D data points. Answer: The first class is correlation based, and the second is feature based. The correlation based methods produce more 3D data points than the feature based methods.
9) Define in one or two sentences the epipolar plane, the epipole and the epipolar constraint. Instead of using English sentences you can draw a single picture to define both the epipolar plane and the epipoles. However, you must use an English sentence to define the epipolar constraint. Answer: Assume P is a 3D point in a stereo system that is seen by the two cameras. The epipolar plane is the plane through P and the center of projections of both cameras. The epipole is the image in one camera of the center of projection of the other camera. The epipolar constraint says that corresponding image points between two cameras lie on the same line. 1 ½ marks

10) In a general stereo system there is a rotation R, and a translation T between the cameras. The equation of a general skew symmetric matrix is defined on the last page. Write the equation of the essential matrix in terms of the rotation R, and the elements of the translation vector T. Answer: The essential matrix E = R S. Here R is the 3 by 3 rotation matrix and S is the 3 by 3 skew symmetric matrix as shown in my diagram with T = (Tx, Ty, Tz) in place of (a1, a2, a3). 1 mark

11) There is a stereo system and point in 3D space which projects to points $\mathbf{p_L}$ and $\mathbf{p_R}$ in camera co-ordinates, and this same point has pixel co-ordinates of $\boldsymbol{p_L}$ and $\boldsymbol{p_R}$. Assume the camera calibration matrix K is the same for both stereo cameras. First write down two equations, one that relates $\mathbf{p_L}$, K and $\boldsymbol{p_L}$, and another that relates $\mathbf{p_R}$, K and $\boldsymbol{p_R}$. Given an essential matrix E, it is the case that $\mathbf{p_L}^{\mathbf{T}}$ E $\mathbf{p_R}$ = 0. Use this relationship, along with the first two equations that you wrote down, to derive a similar relationship between the fundamental matrix F, and pixel co-ordinates $\boldsymbol{p_L}$, and $\boldsymbol{p_R}$. Answer: K is the 3 by 3 camera calibration matrix of both cameras. K has rank 3 and is invertible by definition. Then $\boldsymbol{p_L} = K\,\mathbf{p_L}$, $\boldsymbol{p_R} = K\,\mathbf{p_R}$ and $\mathbf{p_L} = K^{-1}\boldsymbol{p_L}$, $\mathbf{p_R} = K^{-1}\boldsymbol{p_R}$.. We know that $\mathbf{p_L}^{\mathbf{T}}$ E $\mathbf{p_R}$ = 0 so this means that $(K^{-1}\boldsymbol{p_L})^T E\, K^{-1}\boldsymbol{p_R} = 0$. By the characteristics of the transpose operation T, this means that $\boldsymbol{p_L}^T K^{-T} E\, K^{-1}\boldsymbol{p_R} = 0$. The quantity $K^{-T} E\, K^{-1}$ is known as the fundamental matrix F. 1 ½ marks

12) Mark each of the following sentences as either true or false: a) the Fundamental matrix has rank 2. b) the Essential matrix has rank 2. c) every epipolar line goes through the eipole d) the epipole is always visible in the image plane e) the epipolar lines are always parallel in a simple stereo configuration. Answer: a) True b) True – both essential and fundamental matrix have rank 2 c) True – every epipolar line goes through the epipole d) False – the epipole is not necessarily in the camera plane (example, two images that have little or no rotation) e) True. 1 ¼ marks

13) What is the minimum number of corresponding points between two images that are necessary to compute the fundamental matrix F. Why is the minimum number necessary? Answer: The fundamental matrix has 9 elements and therefore has nine degrees of freedom but the system is homogeneous so there are actually only 8 degrees of freedom. Each correspondence sets one degree of freedom, so we need at least 8 correspondences to compute the fundamental matrix. 1 mark

14) When only the intrinsic parameters of a stereo system are known the reconstruction can be computed only up to a scale factor. Explain why this is the case in one or two sentences. Answer: If we do not know the intrinsic parameters we do not know the actual length of the baseline, so we must decompose the essential matrix E into an R, and T. But the scale of translation can not be known because if the triangle of point P, and the camera centers of projection were scaled up then the same correspondences would exist. So we can not know the actual translation, or even the actual depth, but we can know it up to a scale factor (P.S. I accepted simpler answers than this). ½ mark

15) A given 3D point **X** has a projection in pixel co-ordinates of **x**. The projection equation for the camera at the origin is $\mathbf{x} = K\ \mathbf{X}$ and if the camera is rotated by the rotation matrix R without any translation the same 3D point X projects to a new pixel co-ordinate **x'** using the equation $\mathbf{x'} = K\ R\ \mathbf{X}$. Derive a direct relationship between **x** and **x'**. Answer: In these equations the 3D point X is the same, it is only x and x' that change. So since K is invertable from the first equation it is the case that $X = K^{-1}\ x$ and simply substituting this value for X in the second equations gives us $x' = K\ R\ K^{-1}\ x$. This is a relationship between x and x'. 1 ½ marks

16) What is the purpose of rectification? Answer: The purpose of rectification is to transform a general stereo situation into a simple stereo configuration so that correspondence can be done more efficiently. Remember that in a simple stereo configuration correspondence is easier because epipolar lines are horizontal.
½ marks