Notes for Midterm

Camera

Pinhole:

* As the aperture size increases, the number of light rays that passes through the barrier consequently increases
* Camera lenses have another interesting property: they focus all light rays traveling parallel to the optical axis to one point known as the **focal point.**
* The distance between the focal point and the center of the lens is commonly referred to as the **focal length** f.
* We classify the radial distortion as pincushion distortion when the magnification increases and barrel distortion when the magnification decreases. Radial distortion is caused by the fact that different portions of the lens have differing focal lengths.

The following are captured in the camera matrix:

* Pixel size
* Lens Distortion
* Origin of the pixels in the image
* Focal length

Camera Matrix Model:

* If k = l, we often say that the camera has square pixels.
* A point in 3D can be related to its image by the following equation P’ = MP, which is also the same as P’ = K[I 0] P. K here is referred to as the camera matrix.
* We discount distortion for the class so our camera matrix had 5 degrees of freedom, 2 for focal length, 2 for offset and 1 for skewness.
* Intrinsic parameters map from the camera reference system to the image reference system.
* Extrinsic parameters map from the world reference system to the camera reference system. So for extrinsic case to map from the world reference system to the camera reference system we use the following equation P’ = K [R T] Pw = M Pw where full projection matrix M contains the intrinsic and extrinsic parameters.
* Overall the camera matrix now has 11 degrees of freedom.

**Camera Calibration:**

Estimating the camera matrix parameters from the images.

* Each correspondence point gives us 2 equations to solve for unknowns. Therefore we need at least 6 points to solve for the 11 unknowns.
* Not all sets of n correspondences will work. For example, if the points Pi lie on the same plane, then the system will not be able to be solved. These unsolvable configurations of points are known as degenerate configurations.

Handling Distortion in Camera Calibration:

* Often, distortions are radially symmetric because of the physical symmetry of the lens.

Rigid Transformations:

The basic rigid transformations are rotation, translation and scaling.

Rotation

* Rotation matrices are square orthogonal matrices with a determinant of 1.
* We always rotate in the counter-clockwise direction.

Translation:

* Translation can define movement in a certain direction
* Projective transformations occur when the final row of T is not [0001]

Graphical user interface, diagram, application

Description automatically generated with medium confidence

Single View Meteorology

* Isometric transformations are transformations that preserve distances. For example rotation and translation.
* Similarity transformations are transformations that preserve shape. They can do everything that isometric transformations do plus scaling.
* Affine transformations are transformations that preserve points, straight lines and parallelism. T(v) = Av + t
* Projective transformation or hymnographies are any transformation that maps lines to lines but does not necessarily preserve parallelism.
  + It does preserve collinearity of points.
  + The cross ratio of 4 collinear points is invariant.

Points and Lines at infinity:

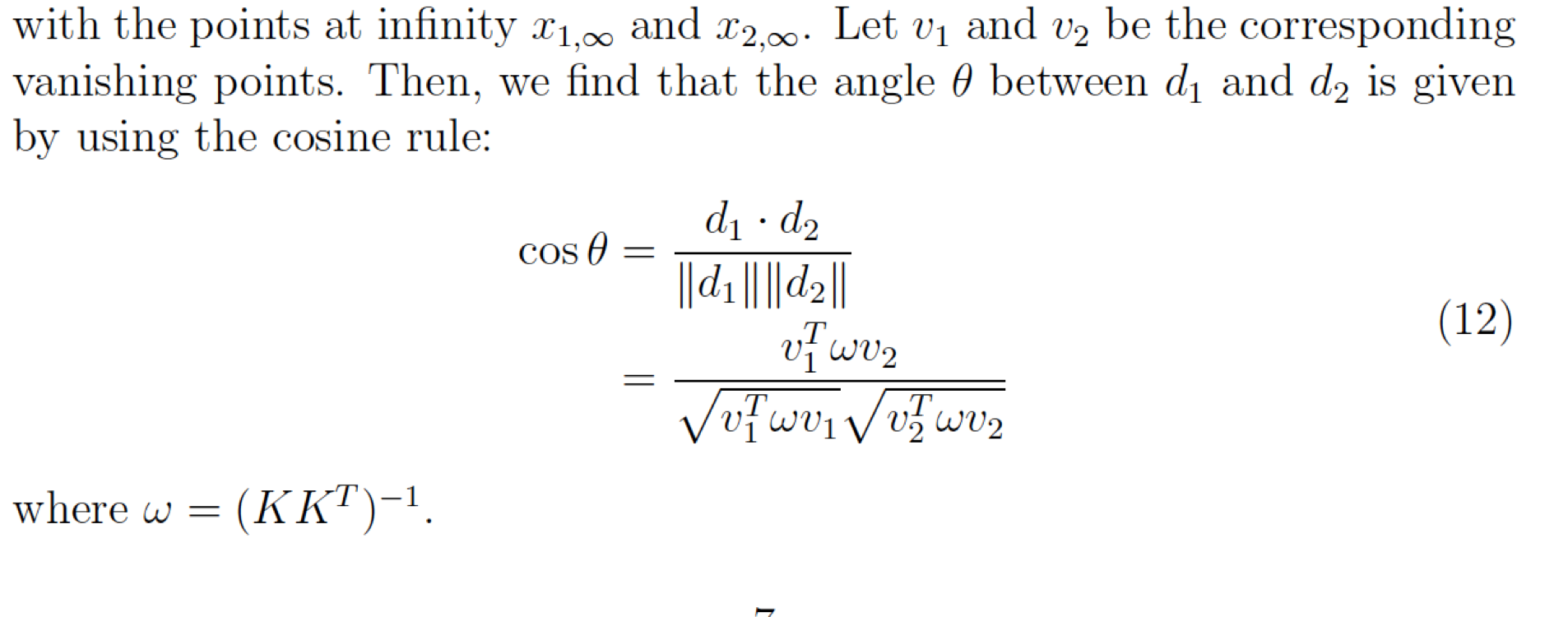
* The intersection point between 2 lines is the cross product of the 2 lines.
* Projective transformations of lines at infinity does not necessarily map to another line at infinity.
* Affine transformations of lines at infinity will map to another line at infinity.

A picture containing text

Description automatically generated

Vanishing points and lines:

* Lines in 3D are the intersection of 2 planes
* Points at infinity in 3D are the intersection of parallel lines in 3D.
* if we apply a projective transformation to one of these points at infinity x1, then we obtain a point p1 in the image plane, which is no longer at infinity in homogeneous coordinates. This point p1 is known as a vanishing point.
* d = (a; b; c) is the direction of a set of 3D parallel lines in the camera reference system. Then vanishing point is v = Kd where K is the intrinsic camera matrix. d = inv(K)v/||inv(K)v||
* The projective transformation of a line at infinity to the image plane is the horizon line
* Normal n of a plane in 3D n = K.T l\_horizon.



Text

Description automatically generated

A single image can be used to get the camera matrix if we know 3 vanishing points and there is no skew and square pixels in the camera.

Epipolar Geometry

In general, it is not possible to recover the entire structure of the 3D world from just one image due to the intrinsic ambiguity of 3D to 2D mapping, some information is simply lost.

The geometry that relates the cameras, points in 3D, and the corresponding observations is referred to as the epipolar geometry of a stereo pair.

* The line between the 2 camera centers is known as the **baseline.**
* The plane defined by the 2 camera centers and P is known as the **epipolar plane.**
* The location of where the baseline intersects the 2 image planes is known as the **epipoles** e and e’.
* The lines defined by the intersection of the epipolar plane and the 2 image planes are known as **epipolar lines.**
* When 2 image planes are parallel then the epipoles are located at infinity. Since the baseline joining the 2 centers is parallel to the image planes.
* Epipolar lines are parallel to an axis of the image plane

**The Essential Matrix**

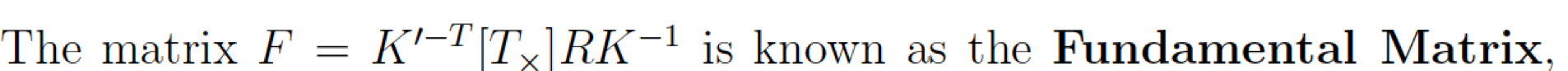
* Assuming we have canonical camera where K = K’ = I then we have the camera matrices as follows:
  + M = [I 0] and M’ = [R.T – R.TT]

Table

Description automatically generated

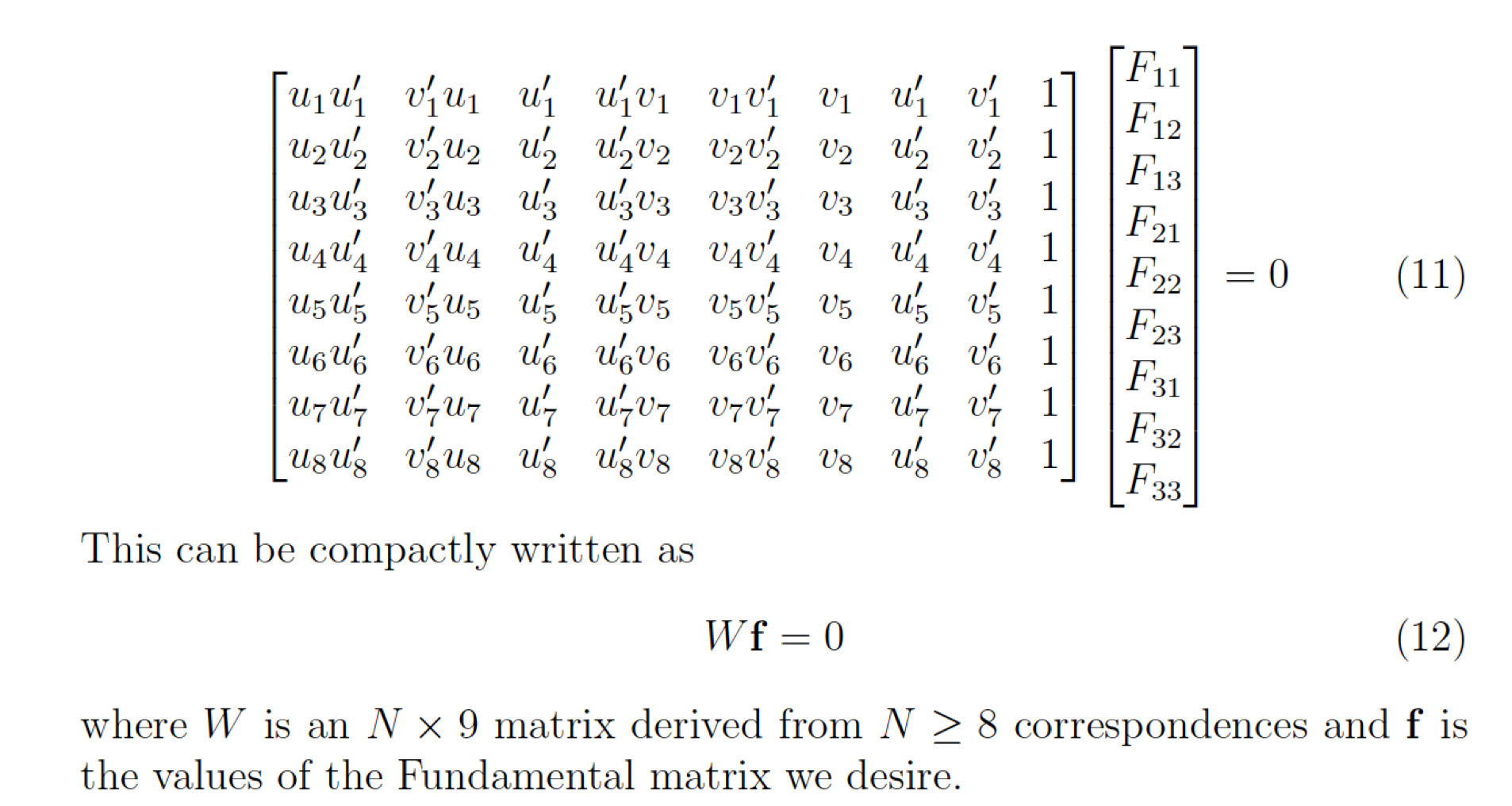
* Essential matrix E = [Tx]R.
* Epipolar constraint p.T E p’= 0
* The essential matrix is a 3x3 matrix with 5 degrees of freedom. It has rank = 2 and is singular.
* To get epipolar lines l’ = E.T p and l = Ep’
* E.T e = 0 and E e’ = 0

**The fundamental Matrix**



It acts similar as the essential matrix but it also encodes information about the camera matrices and the relative rotation and translation between cameras.

* Contains 7 degrees of freedom



Then we can compute F using the 8 points algorithm.

Image Rectification

Rectification is the process of making any 2 given images parallel and is useful for discerning relationships between corresponding points.

To rectify images:

* Get the fundamental matrix
* Get the epipolar lines
* Estimate the epipoles from the epipolar lines
* In real world due to noisy measurements all the epipolar lines will not intersect to a single point.
* After finding the epipoles we will notice that they are not points at infinity along the horizontal axis.
* This means that we need to find a pair of homographies that we can apply to the images to map the epipoles to infinity.
* See pages 12 and 13 of epipolar geometry for how to compute H1 and H2.

Stereo Systems and Structure From Motion

This focuses on how to recover information of the 3D scene from multiple 2D images.

Triangulation: The process of determining the location of a 3D point given its projection into 2 or more images

* Reprojection error: The reproject error for a 3D point in an image is the distance between the projection of the point in the image and the corresponding observed point in the image plane

Structure from motion

* Combining multiple views to simultaneously determine the 3D structure of the scene and the camera parameters.
* Structure from motion problem has 8m + 3n unknowns

The factorization method:

* Center the image points.
* Solve the equation x = AX to get the motion and structure matrices.
* Put the image points into D which will be a 2mxn matrix of rank 3
* The Motion matrix are the non zero values of u after SVD which is the top left 3x3 matrix. It is also 2mx3
* The structure matrix is the diagonal matrix of singular values S dotted with the vt matrix top3 rows. It is 3xn

Ambiguities:

* Metric reconstruction, a reconstruction with similarity ambiguity

Perspective Structure From Motion:

* In the general case with projective cameras, each camera Matrix Mi contains 11 degrees of freedom and is defined up to a scale.

Table

Description automatically generated

Because cameras and points can only be recovered up to a 4x4 projective transformation up to scale (15 parameters), we have 11m + 3n - 15 unknowns in 2mn equations. From these facts, we can determine the number of views and observations that are required to solve for the unknowns.

Determining motion from the Essential Matrix:

* Relationship between the Essential Matrix and the Fundamental Matrix E = K.T F K
* Essential Matrix assumes that we have calibrated cameras and therefore we know the intrinsic camera parameters.
* Essential Matrix only has 5 degrees of freedom because it only encodes the extrinsic parameters.

**Bundle Adjustment:**

Limitations of the previous methods:

* The Factorization method assumes that all points are visible in every image.
* Finally the algebraic approach produces pairwise solutions that can be combined into a camera chain, but does not solve for a coherent optimized reconstruction using all the cameras and 3D points

To resolve these problems we introduce bundle adjustments which is a non-linear method of solving the structure from motion problem.

* Bundle adjustment handles several cameras, it only calculates the reprojection error for only the observations that can be seen by each camera.
* it can handle a large number of views smoothly
* Can handle cases when particular points are not observable by every image

Limitation is that it is a large minimization problem.

Active and Volumetric Stereo

Active stereo helps to mitigate the correspondence problem from the traditional stereo.

The main idea of active stereo is to replace one of the two cameras with a device that interacts with the 3D environment, usually by projecting a pattern onto the object that is easily identifiable from the second camera. This new projector-camera pair defines the same epipolar geometry that we introduced for camera pairs, whereby the image plane of the replaced camera is replaced with a projector virtual plane

Volumetric Stereo:

* We assume that the 3D points we are trying to estimate is within some contained known volume.
* Has 3 different methods: Space carving, shadow carving and voxel coloring.

Space carving:

* Stems from the observation that the contours of an object provide a rich source of geometric information about the object.
* Each camera observes some visible portion of the object from which contours can be gotten.
* When projected into the image plane, this contour encloses a set of pixels known as the silhouette of the object in the image plane.
* Space carving then uses the silhouettes of objects from multiple views to enforce consistency.
* Visual cone: The enveloping surface defined by the camera center and the object contour in the image plane. By construction it is guaranteed that the object will lie completely in both the initial volume and the visual cone.
* Visual Hull: Intersection of visual cones from each camera.

Algo:

* For example, if our cameras encircle the object, then we can simply say that the working volume is the entire interior of the space enclosed by the cameras.
* We divide this volume into small units known as voxels, defining what is known as a voxel grid.
* We take each voxel in the voxel grid and project it into each of the views. If the voxel is not contained by the silhouette in a view, then it is discarded.
* Consequently, at the end of the space carving algorithm, we are left with the voxels that are contained within the visual hull.

Limitations:

* Scales linearly with the number of voxels in the grid.
* Another limitation is that the efficacy of space carving is dependent on the number of views, if the number of views is too low then we end up with a very loose estimate of the visual hull of the object.
* the preciseness of the silhouette. If too conservative and contains more pixels than necessary then it will be imprecise.
* and even the shape of the object we are trying to reconstruct.
* Incapable of rendering certain concavities of the object.

SHADOW CARVING

* Self-shadows are the shadows that an object projects on itself

VOXEL COLORING

Uses color consistency instead of contour consistency.

* For each voxel, we look at its corresponding projections in each of the images and compare the color of each of these projections.
* If the colors of these projections sufficiently match, then we mark the voxel as part of the object.
* One benefit of voxel coloring not present in space carving is that color associated with the projections can be transferred to the voxel, giving a colored reconstruction.
* Critical assumption for color consistency is that it must be Lambertian which means that the perceived luminance of any part of the object does not change with viewpoint, location or pose.
* Voxel coloring produces a solution that is not unique.

Chart

Description automatically generated