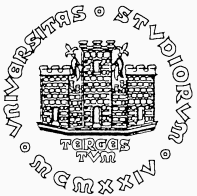
Computer Vision and Pattern Recognition

Course ID: 554SM – Fall 2018

Felice Andrea Pellegrino

University of Trieste
Department of Engineering and Architecture



554SM –Fall 2018

Lecture 6: Stereopsis

Triangulation

Human stereopsis

[...] the two eyes form slightly different images of the world. The relative difference in the positions of objects in the two images is called disparity, which is caused by the differences in their distance from the viewer. Our brains are capable of measuring this disparity and of using it to estimate the relative distance of the objects from the viewer.

(David Marr)

Computational stereopsis

The goal of *computational* stereopsis is to get information about the three dimensional structure of a scene, from a pair of images captured by two cameras in different positions.

We can distinguish two steps/subproblems:

- computing the correspondences;
- performing triangulation.

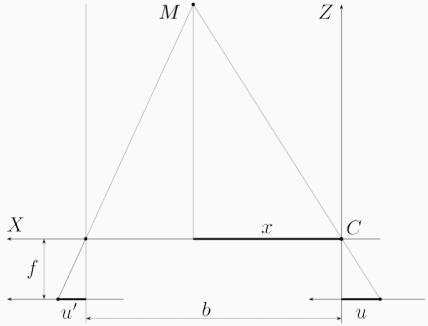
The first is based on feature detection and matching. Two points in the two images that correspond to the same point of the scene are called *conjugate points*.

Given the location of the conjugate points corresponding to the 3D point M of the scene, and the two perspective projection matrices (i.e. the intrinsic and extrinsic parameters of the two cameras), the location in the 3D space of M can be found by *triangulation*.

Triangulation (normal case)

- Consider two cameras, aligned and having parallel optical axes.
- The disparity is then purely horizontal, thus the 2D model in figure applies.
- Placing the world reference frame in correspondence of the camera on the right, the perspective projection equations are:

$$\begin{cases} \frac{f}{z} = \frac{-u}{x} \\ \frac{f}{z} = \frac{-u'}{x-b} \end{cases}$$



- solving for z we get

$$z = \frac{bf}{u' - u}.$$

  - Thus, from the disparity $u' - u$ and the geometry of the stereo pair (the *baseline* b and f) it is possible to compute the depth of point M.
  - Notice that b acts as a scale factor and that the depth z is inversely proportional to the disparity.

Observation (sensitivity to disparity error)

Suppose that the disparity is affected by an error Δd. What is the error on the computed depth z?

By taking the derivative of $z = \frac{bf}{d}$ w.r.t. d we get:

$$\frac{\partial z}{\partial d} = -\frac{bf}{d^2}.$$

thus, for small errors Δd we get

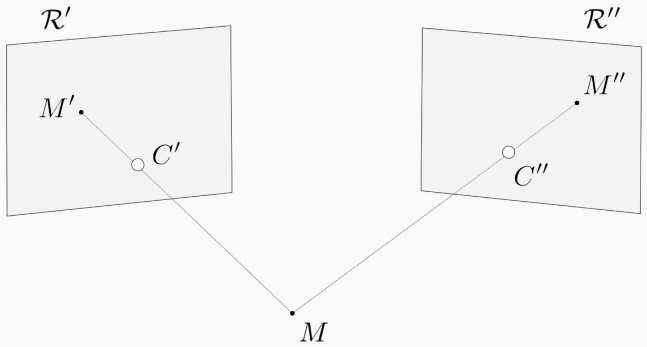
$$\Delta z \approx -\frac{bf}{d^2}\Delta d.$$

By substituting the expression for z, we obtain

$$\Delta z \approx -\frac{z^2}{bf}\Delta d.$$

As a consequence, for a given Δd, the error is proportional the square of the working distance z and inversely proportional to the baseline b and to the focal length f.

Triangulation (general case)



- Let P' and P'' be the two perspective projection matrices and let $m' = [u'\ v'\ 1]^\top$ be the homogeneous coordinates of M' and $m'' = [u''\ v''\ 1]^\top$ be the homogeneous coordinates of M''. Let $m = [x\ y\ z\ 1]^\top$.
- As for M', we get

$$\begin{cases} (p'_1 - u' p'_3)^\top m = 0 \\ (p'_2 - v' p'_3)^\top m = 0 \end{cases}$$

Triangulation (general case) (cont.)

- By considering the conjugate point M'', we get another pair of equations, thus we can write a homogeneous linear system of 4 equations and 4 unknowns:

$$\begin{bmatrix} (p'_1 - u' p'_3)^\top \\ (p'_2 - v' p'_3)^\top \\ (p''_1 - u' p''_3)^\top \\ (p''_2 - v' p''_3)^\top \end{bmatrix} m = 0.$$

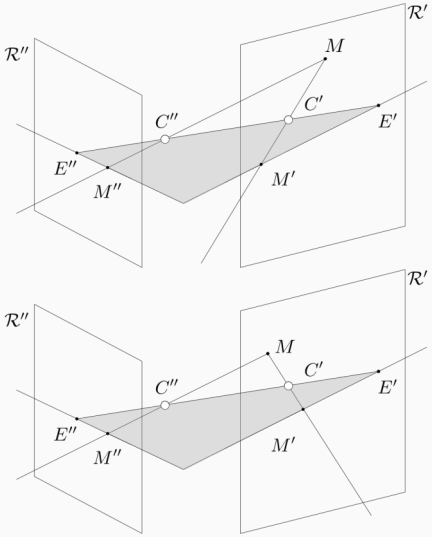
- The solution is the null space of the 4×4 matrix of coefficients. The matrix must be rank 3 otherwise the only solution is the trivial solution $m = 0$.
- In practice, due to noise, the rank is 4 and a least squares solution is sought for, using singular value decomposition.
- The method can be generalized to $N > 2$ cameras, leading to $2N$ linear equations.
- The corresponding geometric residual (that leads to better results, but requires solving a non linear minimization problem) is

$$\epsilon(m) = \sum_{j=1}^{N} \left(\frac{p_1^{j\top} m}{p_3^{j\top} m} - u^j \right)^2 + \left(\frac{p_2^{j\top} m}{p_3^{j\top} m} - v^j \right)^2.$$

Epipolar geometry

Epipolar lines

- Epipolar geometry describes the relationship between conjugate points.
- With reference to the figures, given M'', any possible conjugate point M' must lie on the intersection between $\mathcal{R}'$ and the plane containing M, C' and C''.
- Such intersection is a straight line called *epipolar line* (relative to M'').
- All the epipolar lines in $\mathcal{R}'$ pass through the point E' (the *epipole*).
- Similarly, all the epipolar lines in $\mathcal{R}''$ pass through the epipole E''.
- The plane containing M, C' and C'' is the *epipolar plane* of M.
- E' is the projection of C'' in $\mathcal{R}'$ and similarly, E'' is the projection of C' in $\mathcal{R}''$.



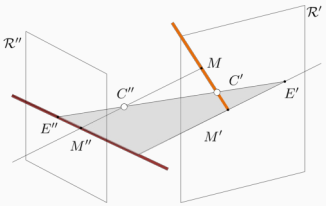
Example

A stereo pair. Epipolar lines corresponding to the points on the left image, are drawn on the right image.

Equation of the epipolar line

- Let $P' = [Q'|q']$ and $P'' = [Q''|q'']$. Then

$$\begin{cases} m' \simeq P'm \\ m'' \simeq P''m \end{cases}$$



- The epipolar line corresponding to M' (violet) is the projection through P'' of the optical ray of M' (orange), whose equation is

$$m = c' + \lambda \begin{bmatrix} (Q')^{-1}m' \\ 0 \end{bmatrix}.$$

  - Thus, by substitution:

$$m'' \simeq P''m = P'' \left(c' + \lambda \begin{bmatrix} (Q')^{-1}m' \\ 0 \end{bmatrix} \right) = \underbrace{P''c'}_{e''} + \lambda \left[Q'' \,|\, q'' \right] \begin{bmatrix} (Q')^{-1}m' \\ 0 \end{bmatrix} =$$

$$e'' + \lambda Q''(Q')^{-1}m'.$$

Equation of the epipolar line (cont.)

- Thus, the epipolar line of M' has the following equation:

$$m'' \simeq e'' + \lambda Q''(Q')^{-1} m'.$$

- Now, multiplying both terms by $[e'']_\times$ we get

$$[e'']_\times m'' \simeq [e'']_\times (e'' + \lambda Q''(Q')^{-1} m') = \lambda [e'']_\times Q''(Q')^{-1} m'.$$

- Finally, since the term on the left side is orthogonal to m'', if we multiply both terms by $m''^\top$ we get

$$0 = m''^\top [e'']_\times Q''(Q')^{-1} m',$$

  which is called the *Longuet-Higgins equation* and is a bilinear form.

- By letting $F = [e'']_\times Q''(Q')^{-1}$, the Longuet-Higgins equation becomes

$$m''^\top F m' = 0.$$

- The matrix F is called the *fundamental matrix*. Given m', the triplet $\lambda = Fm'$ defines its epipolar line by the equation

$$\lambda^\top m'' = 0.$$

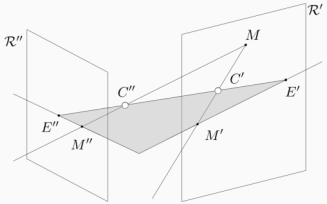
The fundamental matrix

Facts about the fundamental matrix:

- F defines **a relationship between (homogeneous) pixel coordinates of conjugate points** in a pair of images;
- F **has rank 2**, because
$$F = \left[e''\right]_\times Q''(Q')^{-1}$$
thus it is the product of a skew symmetric matrix (having rank 2) and the two full rank matrices Q'' and $(Q')^{-1}$;
- F **encodes information on both the intrinsic and the extrinsic parameters** (because Q' and Q'' depend on both);
- it can be **computed from the two projection matrices** P' and P'';
- it can be **estimated from point matches in pixel coordinates** only (as we will see later on).

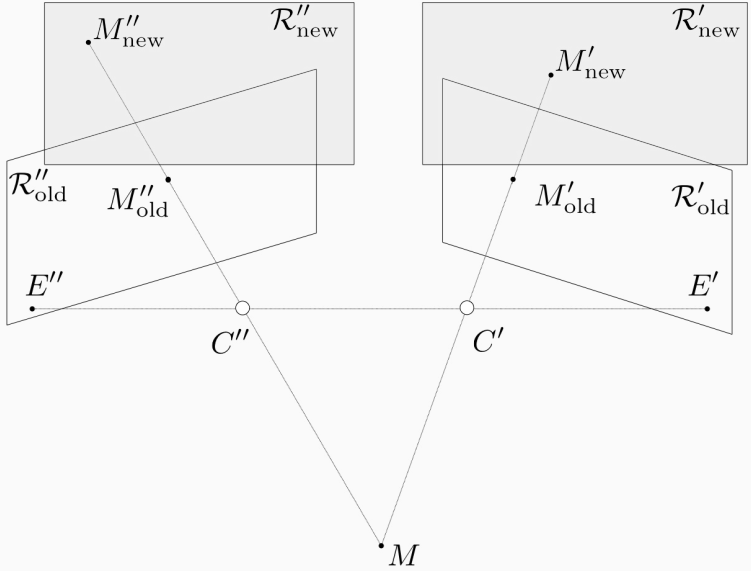
Epipolar rectification

- If C' is in the focal plane of the conjugate camera (the plane parallel to $\mathcal{R}''$ and containing C''), then E'' projects to infinity and **the epipolar lines are parallel**.
- When both the epipoles are located at infinite, the epipolar lines are parallel in both the images (for this to happen, the baseline $C'C''$ must be contained in both the focal planes, or equivalently, the focal planes must be coincident).

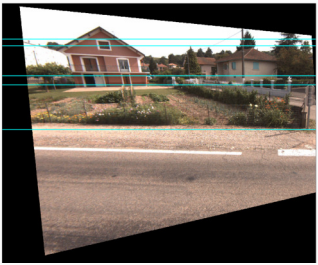


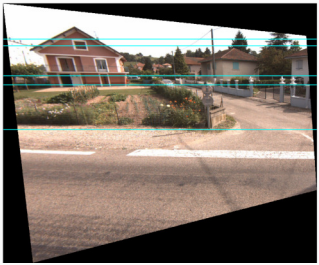
  - When the epipolar lines are parallel and horizontal, the correspondence problem if simplified in that **the potential matches lie on the same image row**.
  - The *epipolar rectification* is the process of recovering that favorable configuration by means of suitable transformations.
  - The idea is to **define a new pair of perspective projection matrices**, with the same points of view C' and C'' but rotated to get $\mathcal{R}'$ and $\mathcal{R}''$ parallel to the baseline.

Epipolar rectification (cont.)



Epipolar rectification (cont.)





Epipolar rectification (cont.)

Procedure for epipolar rectification, from (Fusiello et al., 2000).

Suppose that the two perspective projection matrices (PPMs) P'_{old} and P''_{old} are known. We define two new PPMs P'_{new} and P''_{new} obtained by:

1. rotating the old ones around their optical centers until **focal planes becomes coplanar** (i.e. they contain the baseline $C'\,C''$); thus the epipoles are at infinity and the epipolar lines are parallel;
2. choosing, among the possible rotations that guarantee the previous condition, a pair of rotations such that **the new X axes of both cameras are parallel to the baseline**; this ensures that the epipolar lines are horizontal;
3. requiring that **the new cameras have the same intrinsic parameters**; this ensures that conjugate points have the same vertical coordinate.

Note that the previous conditions guarantee that the retinal planes $\mathcal{R}'_{\text{new}}$ and $\mathcal{R}''_{\text{new}}$ are coplanar.

In summary, the new cameras will have the same intrinsic parameters (same K), the same orientation of the camera frame (same R) but will differ for the centers of projection (that will be the same centers of the old cameras).

Epipolar rectification (cont.)

Recalling that a PPM can be factorized as

$$P = [Q \mid q] = [KR \mid Kt] = K[R \mid t]$$

and that the Cartesian coordinates $\tilde{c}$ of the center C may be expressed as

$$\tilde{c} = -Q^{-1}q, \tag{1}$$

it follows that

$$q = -Q\tilde{c} = -KR\tilde{c}.$$

Thus, the new PPMs may be factorized as

$$P'_{\text{new}} = K[R \mid -R\tilde{c}'] \qquad \text{and} \qquad P''_{\text{new}} = K[R \mid -R\tilde{c}''].$$

The intrinsic parameter matrix K can be choosen arbitrarily. The centers $\tilde{c}'$ and $\tilde{c}''$ are the same of the old cameras and can be computed from the PPMs using equation (1). Thus, we only need to choose an appropriate R.

Epipolar rectification (cont.)

As for the rotation matrix R, observe that its rows

$$R = \begin{bmatrix} r_1^\top \\ r_2^\top \\ r_3^\top \end{bmatrix},$$

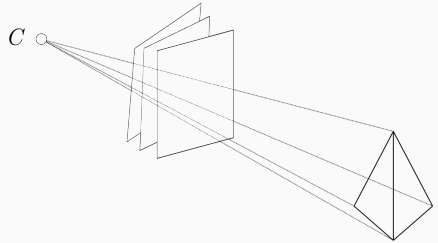
represent the axes X, Y and Z of the camera reference frame w.r.t. the world reference frame. Thus, to fulfill the previous requirements, they can be chosen as follows.

1. $r_1 = (\breve{c}'' - \breve{c}') \,/\, \|\breve{c}'' - \breve{c}'\|$, to get a new X axis parallel to the baseline;
2. $r_2 = k \times r_1$, to get a new Y axis orthogonal to X and to an arbitrary unit vector k;
3. $r_3 = r_1 \times r_2$, to get a new Z axis orthogonal to both X and Y.

A possible choice in step 2 is taking k equal to the Z vector of an old camera.

The algorithm fails when the optical axis is parallel to the baseline, i.e., when there is a pure forward motion.

Epipolar rectification (cont.)



Given the new PPMs we can **rectify the images**, i.e. recovering the images that the new cameras would capture, from the images of the old cameras.

Suppose we want to rectify the first camera: we need to compute the transformation mapping the image plane of P'_{old} onto the image plane of P'_{new}.

The optical center is the same, thus the situation is illustrated in the figure, where, moving the image plane we get different images of the same scene.

The sought transformation turns out to be a **homography**.

Epipolar rectification (cont.)

Indeed, for a generic point M we can write

$$\begin{cases} m'_{\text{old}} \simeq P_{\text{old}} m \\ m'_{\text{new}} \simeq P_{\text{new}} m \end{cases}$$

where we use the prime ($'$) to denote image coordinates. On the other hand, the equations of the optical rays of M are the following (since rectification does not move the optical center):

$$\begin{cases} \tilde{m} = \tilde{c} + \lambda_{\text{old}} Q_{\text{old}}^{-1} m'_{\text{old}} \\ \tilde{m} = \tilde{c} + \lambda_{\text{new}} Q_{\text{new}}^{-1} m'_{\text{new}} \end{cases}.$$

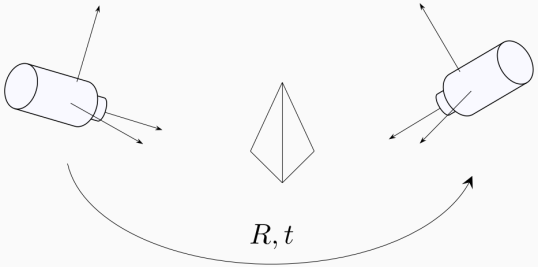
Thus, by subtracting we get

$$m'_{\text{new}} \simeq Q_{\text{new}} Q_{\text{old}}^{-1} m'_{\text{old}},$$

hence the mapping is a homography defined by $T = Q_{\text{new}} Q_{\text{old}}^{-1}$.

Relative pose

Relative pose

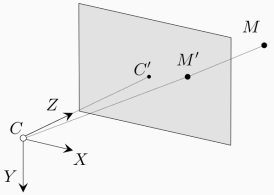


- Suppose that the same scene is captured by two cameras (or by the same camera in two different positions/orientations).
- The *relative pose* problem is that of **determining the pose of one camera w.r.t. the other, assuming that the intrinsic parameters of both are known**.
- When no further information is available, one of the two cameras is set at the origin of the world reference frame and at a canonical orientation, thus having a PPM $P = K[I\ 0]$.
- As a consequence, the relative pose problem is that of finding the extrinsic parameters R, t of the other camera.

Normalized coordinates

Take a point M and let $m = \lambda[x\ y\ z\ 1]^\top$ be a vector of homogeneous coordinates. Applying the perspective projection we get

$$m' = Pm = K[R \quad t]m.$$



Consider a new vector of coordinates defined as

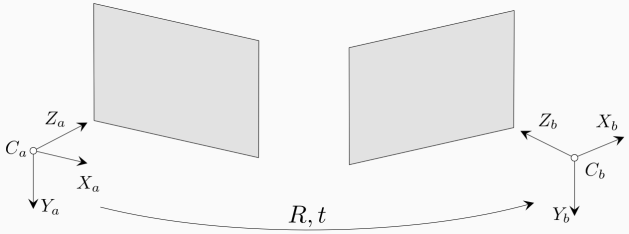
$$\underline{m}' = K^{-1}m,$$

which basically inverts the last transformation (notice that K is invertible by definition of PPM). Then $\underline{m}'$ is the vector of *normalized coordinates*.

It is a vector of homogeneous coordinates of the image points, expressed in units of focal length, with respect to the principal point. For instance, $\underline{m}' = [2\ 1\ 1]^\top$ refers to the point $[2f\ f]$ w.r.t. the principal point (C' in the figure), or the 3D point $[2f\ f\ f]^\top$ w.r.t. the camera reference frame.

If $\lambda = 1$, $\underline{m}'$ can be regarded as a vector of Cartesian coordinates, representing M w.r.t. the camera reference frame. For positive λ, it has **the same direction of the optical ray** through M.

Epipolar constraint



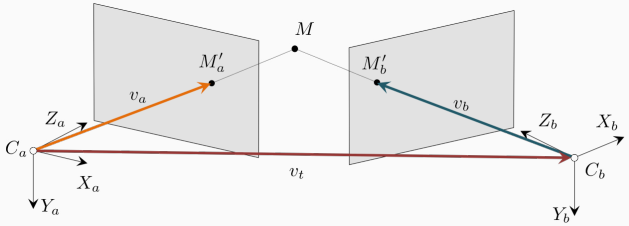
- Suppose we have two cameras a and b, whose PPMs are respectively

$$P_a = K_a[I \quad 0] \qquad \text{and} \qquad P_b = K_b[R \quad t].$$

- The particular form of P_a tells that the world reference frame is equal to the camera reference frame (X_a, Y_a, Z_a, C_a) of camera a).
- The pair R, t represents the rigid transformation (i.e. the relative pose) from the world reference frame to that of camera b. Thus, it also represents the transformation from camera a to camera b.
- Suppose that K_a and K_b are known (the intrinsic parameters of both cameras are known). As a consequence, we can recover the normalized coordinates from the image coordinates by inverting K_a and K_b:

$$\underline{m}'_a = K_a^{-1} m'_a \qquad \text{and} \qquad \underline{m}'_b = K_b^{-1} m'_b.$$

Epipolar constraint



Consider the vectors v_a, v_b and v_t in figure. By construction, they belong to the same plane (the epipolar plane of M), thus they are coplanar. As a consequence, their triple product is zero:

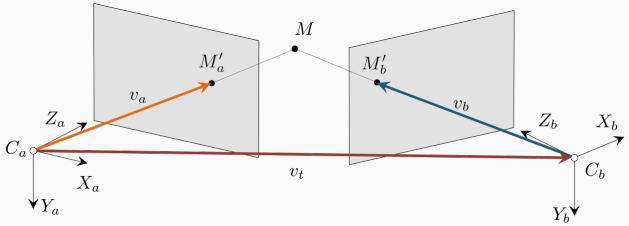
$$v_b \cdot (v_t \times v_a) = 0.$$

Now we represent the three vectors in the reference frame of camera b (we denote the reference frame as a left superscript):

$${}^b v_b = \mu \underline{m}'_b, \qquad {}^b v_t = t, \qquad {}^b v_a = R\nu \underline{m}'_a$$

for some positive μ, ν (we are regarding the normalized coordinates as Cartesian coordinates w.r.t. the camera reference frame).

Epipolar constraint



Thus the condition on the triple product can be expressed as a matrix product as follows

$$\underline{m}_b'^{\top} \underbrace{[t]_\times R}_{\doteq E} \underline{m}'_a = 0,$$

where the matrix E is called the *essential matrix* (Longuet-Higgins, 1981). The equation

$$\underline{m}_b'^{\top} E \underline{m}'_a = 0 \tag{2}$$

is called the *epipolar constraint* and **expresses a constraint on the normalized coordinates of conjugate points as a function of the relative pose of the two cameras**.

The essential matrix

The essential matrix plays a crucial role in the relative pose problem as well as in the *structure from motion* problem (a repeated pose estimation and triangulation that allows to recover a 3D structure from multiple images).

Facts about the essential matrix:

- $E = [t]_\times R$ defines **a relationship between normalized coordinates of conjugate points** in a pair of images;
- E **has rank 2** for any nonzero t because it is the product of the skew symmetric matrix $[t]_\times$ (having rank 2) and the full rank matrix R;
- E **encodes information on the extrinsic parameters only**;
- E **is defined up to a scale factor** (since the equation (2) is homogeneous, if E is an essential matrix for a pair of cameras, so is αE for any nonzero α);
- E **has five degrees of freedom**: three for the rotation R and two for the translation t (a degree of freedom is lost due to the arbitrary scale factor);

The essential matrix (cont.)

- E can be estimated from point matches in pixel coordinates, **provided that the intrinsic parameters are known** (and thus we can recover the normalized coordinates).
- if E is an essential matrix for the pair a, b, then $E^\top$ is an essential matrix for the pair b, a; indeed, by transposing equation (2) we get

$$\underline{m}_a'^\top E^\top \underline{m}_b' = 0,$$

  i.e. the epipolar constraint when $P_b = [I \quad 0]$);
- the essential matrix E is related to the fundamental matrix F as follows:

$$\underline{m}_b'^\top E \underline{m}_a' = \underline{m}_b'^\top \overbrace{(K_b^\top K_b^{-\top})}^{I} E \overbrace{(K_a^{-1} K_a)}^{I} \underline{m}_a' = \underbrace{\underline{m}_b'^\top K_b^\top}_{m_b'^\top} \overbrace{K_b^{-\top} E K_a^{-1}}^{F} \underbrace{K_a \underline{m}_a'}_{m_a'}$$

  thus

$$E = K_b^\top F K_a.$$

Factorization of the essential matrix

Suppose that the matrix E is available (for instance, estimated from correspondences). How can we recover the translational and rotational parts t and R? The following theorem provides an answer.

Theorem (Factorization of the essential matrix)

(Huang and Faugeras, 1989) A real 3×3 matrix E can be expressed as a skew-symmetric matrix postmultiplied by a rotation matrix if and only if one of its singular values is zero and the other two are equal.

To prove the theorem we will need the following lemma.

Lemma (Properties of real skew-symmetric matrices)

Let S be a real skew-symmetric matrix. Then

1. *the nonzero eigenvalues of S are imaginary conjugate pairs, i.e. they are $\pm j\lambda_i$, $i = 1 \dots m$ where j is the imaginary unit;*
2. *S is orthogonally similar to its Jordan real form, i.e. there exist U and Λ such that*

$$S = U \Lambda U^\top$$

where U is orthogonal and $\Lambda = \mathsf{diag}\left(\lambda_1 \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \dots, \lambda_m \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, 0, \dots, 0\right)$.

Factorization of the essential matrix (cont.)

Proof. of the Factorization Theorem (adapted from Fusiello (2018)).

($\Leftarrow$) Let $E = UDV^\top$ be the SVD of E, where $D = \mathsf{diag}(\alpha, \alpha, 0)$, $\alpha > 0$. Observe that

$$D = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & \alpha & 0 \\ -\alpha & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{S'} \underbrace{\begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{R'}, \tag{3}$$

where S' is skew-symmetric and R' is a rotation matrix. Thus

$$E = US'R'V^\top = (US'U^\top)(UR'V^\top) \simeq \underbrace{(US'U^\top)}_{S} \underbrace{\mathsf{det}(UV^\top)(UR'V^\top)}_{R}.$$

If we take $S = US'U^\top$ and $R = \mathsf{det}(UV^\top)UR'V^\top$, the sought factorization is $E = SR$. Indeed, it is easy to show that S is skew-symmetric and R is a rotation matrix (the factor $\mathsf{det}(UV^\top)$ guarantees that the determinant of R is one).

Factorization of the essential matrix (cont.)

($\Rightarrow$) Let $E = SR$, where S is skew-symmetric and R is a rotation matrix. Thanks to the previous lemma, we have

$$S = U \begin{bmatrix} 0 & \alpha & 0 \\ -\alpha & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^\top = US'U^\top$$

for some $\alpha > 0$.

From (3) we get $S' = \mathsf{diag}(\alpha, \alpha, 0)R'^\top$, thus:

$$E = U\,\mathsf{diag}(\alpha, \alpha, 0)(R'^\top U^\top R),$$

which is a singular value decomposition. □

Four solutions

Observe that the factorization (3) is not unique. Indeed it easy to recognize that

$$D = S' R' = S'^{\top} R'^{\top},$$

thus two possible factorizations are

$$E = (U S' U^{\top}) \det(UV^{\top})(U R' V^{\top}) = (U S'^{\top} U^{\top}) \det(UV^{\top})(U R'^{\top} V^{\top})$$

Moreover, since E is defined up to a scale factor, and

$$-D = S'^{\top} R' = S' R'^{\top},$$

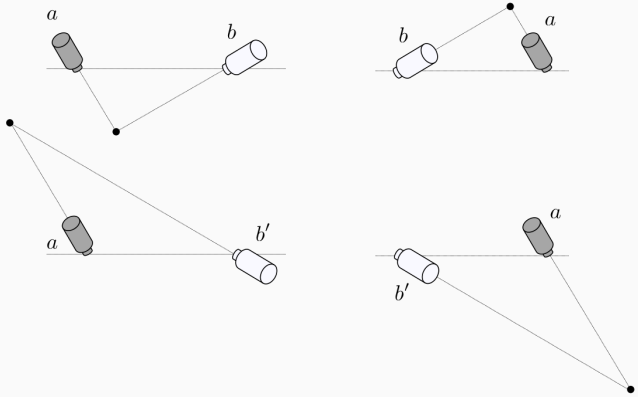
another pair of factorizations is

$$-E = (U S'^{\top} U^{\top}) \det(UV^{\top})(U R' V^{\top}) = (U S' U^{\top}) \det(UV^{\top})(U R'^{\top} V^{\top}).$$

Recalling that $S'^{\top} = -S'$, the four possible S, R pairs are thus:

$$S = \pm U S' U^{\top} \qquad \text{and} \qquad R = \det(UV^{\top}) U \left\{ \begin{array}{c} R' \\ R'^{\top} \end{array} \right\} V^{\top}.$$

Four solutions (cont.)



From a geometrical standpoint, transposing S means changing the verse of t, thus switching the cameras. Transposing R amounts to a rotation of 180 degrees of the camera around the baseline. Only one of the four configurations is admissible (i.e. is such that the points are in front of both the cameras). It can be determined by triangulating one point.

Depth-speed ambiguity

Cameras, fundamentally, measure angles. Thus, the structure (location of the 3D points in the scene) and motion (relative pose of camera b w.r.t. camera a) can be estimated only up to a common nonzero multiplicative scale factor.

Indeed, with no further knowledge than the intrinsic parameters, it is impossible to establish whether the motion observed in the images is due to a close object moving slowly or a distant object moving quickly.

That fact is known as the *depth-speed ambiguity*.

As a consequence, the vector t (the translational component of the motion) can be taken unitary with no loss of information.

The eight-point algorithm

The essential matrix can be estimated from point correspondences by using the *eight-point algorithm* (Longuet-Higgins, 1981) described next.

Given a set of corresponding pairs in normalized coordinates $\{(\underline{m}_a'^{(i)}, \underline{m}_b'^{(i)}), i = 1, \ldots, N\}$, the essential matrix E is sought, such that

$$\underline{m}_b'^{(i)\top} E \underline{m}_a'^{(i)} = 0, \qquad i = 1, \ldots, N.$$

By using the Kronecker product and the vector operator, we get

$$\underline{m}_b'^{(i)\top} E \underline{m}_a'^{(i)} = 0 \quad \Longleftrightarrow \quad \mathsf{vec}\,(\underline{m}_b'^{(i)\top} E \underline{m}_a'^{(i)}) = 0 \quad \Longleftrightarrow \quad (\underline{m}_a'^{(i)\top} \otimes \underline{m}_b'^{(i)\top})\,\mathsf{vec}\,(E) = 0.$$

Thus, for each correspondence, a linear homogeneous equation in 9 unknowns (the entries of E) is obtained. From N correspondences we get a linear homogeneous system:

$$\underbrace{\begin{bmatrix} \underline{m}_a'^{(1)\top} \otimes \underline{m}_b'^{(1)\top} \\ \underline{m}_a'^{(2)\top} \otimes \underline{m}_b'^{(2)\top} \\ \vdots \\ \underline{m}_a'^{(N)\top} \otimes \underline{m}_b'^{(N)\top} \end{bmatrix}}_{A} \mathsf{vec}\,(E) = 0, \tag{4}$$

The eight-point algorithm (cont.)

whose solution is the null-space of A. For $N = 8$, the null-space has dimension one (except for degenerate configurations, see Faugeras and Maybank (1990)), thus the solution is determined up to a multiplicative constant.

In practice, more than eight correspondences are available, and due to noise, the rank of A is 9. Therefore, a least squares solution is found, by solving

$$\underset{||e||=1}{\text{argmin}} \, ||Ae||^2$$

where the constraint forces to discard the trivial solution $e = 0$ (which is not acceptable because $E \neq 0$). The solution of the problem is well-known to be the right singular vector of A corresponding to the smallest singular value. If

$$A = U\Sigma V^\top$$

then $e = v_9$ i.e. the rightmost column of V.

Observations

1. In general, the estimated E does not have a zero singular value and two nonzero equal singular values (meaning that it cannot be factorized as expected). The property can be enforced by substituting $E = UDV^\top$ with $\hat{E} = U\,\mathbf{diag}\,(1, 1, 0)\,V^\top$, which is the closest (in Frobenius norm) matrix to E enjoying the property.
2. In principle, since E has only five degrees of freedom, five correspondences are sufficient for estimating the matrix, provided that we are able to enforce the proper algebraic constraints (basically, the conditions on the singular values). Five-point algorithms do exist, see for instance Li and Hartley (2006).

Algorithm Structure from motion

Input: A set of corresponding points $\{(m_a'^{(i)}, m_b'^{(i)})\}$ in two images; the internal parameters K_a and K_b of the cameras.

Output: The 3D coordinates $m_a^{(i)}$ of the points $M^{(i)}$ (i.e. the structure) and the relative pose R, t of camera b w.r.t. camera a (the motion).

1: Recover the normalized coordinates $\underline{m}_a'^{(i)}$ and $\underline{m}_b'^{(i)}$;
2: compute E using the eight-point algorithm;
3: factorize E is $S = [t]_\times$ and R;
4: instantiate two PPMs $P_a = [I \quad 0]$ and $P_b = [R \quad t]$;
5: perform triangulation of the points and get $m_a^{(i)}$.

Estimating the fundamental matrix

The eight-point algorithm can be used also for the estimation of F.

Given at least eight correspondences of points in **pixel coordinates** $\{(m_a'^{(i)}, m_b'^{(i)}), i = 1, \dots, N\}$, we get a linear homogeneous system in 9 unknowns (the entries of F)

$$\underbrace{\begin{bmatrix} m_a'^{(1)\top} \otimes m_b'^{(1)\top} \\ m_a'^{(2)\top} \otimes m_b'^{(2)\top} \\ \vdots \\ m_a'^{(N)\top} \otimes m_b'^{(N)\top} \end{bmatrix}}_{A} \mathsf{vec}\,(F) = 0, \tag{5}$$

whose solution is the null-space of A.

The obtained $F = U\,\mathsf{diag}(f_1, f_2, f_3)\,V^\top$ is, in general, full rank. The singularity can be enforced by substituting it with $\hat{F} = U\,\mathsf{diag}(f_1, f_2, 0)\,V^\top$.

The seven-point algorithm

However, since F has seven degrees of freedom, it can be estimated by seven correspondences (i.e. seven equations) and the further constraint that $\det(F) = 0$. From seven correspondences we get

$$\underbrace{\begin{bmatrix} {m_a'^{(1)}}^\top \otimes {m_b'^{(1)}}^\top \\ \vdots \\ {m_a'^{(7)}}^\top \otimes {m_b'^{(7)}}^\top \end{bmatrix}}_{A^{7\times 9}} \operatorname{vec}(F) = 0. \tag{6}$$

The matrix $A = UDV^\top$ has a two-dimensional null-space, thus the generic solution can be represented as[1]

$$F = \alpha F_8 + (1-\alpha) F_9,$$

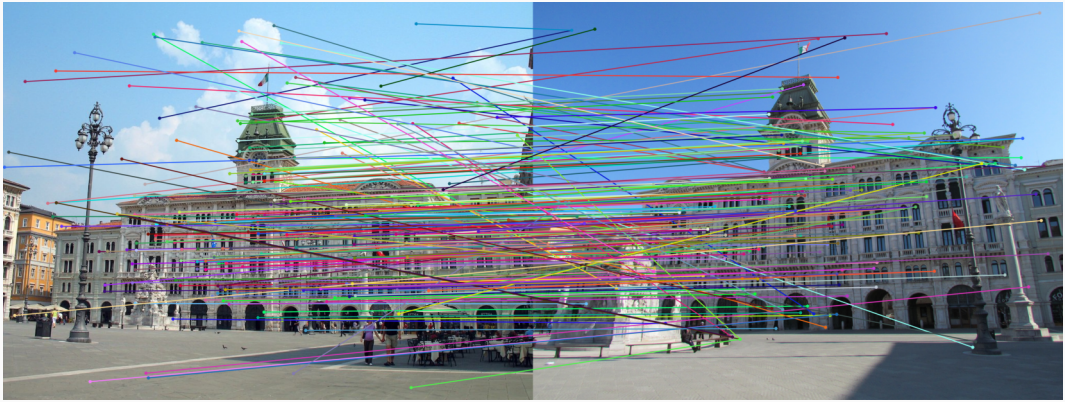
where F_8 and F_9 are the matrices corresponding to the two rightmost columns of V (which span the null-space of A). Thus the singularity constraint becomes

$$\det(\alpha F_8 + (1-\alpha) F_9) = 0,$$

i.e. a polynomial equation of degree three, in α, whose real solutions (one or three) can be found analytically.

[1] Because of its homogeneity, the fundamental matrix is a one-parameter family.

Robust estimation using RANSAC



Correspondences found by applying SIFT to both images and nearest neighbor distance ratio matching.

Robust estimation using RANSAC

Correspondences that are inliers w.r.t. an estimated F, computed using RANSAC. See Lab Lecture for details.

Normalized eight-point algorithm

The eight-point algorithm suffers from poor conditioning (Hartley, 1995). Indeed, the equation corresponding to the pair $m'_a = [u_a \ v_a \ 1]^\top$ and $m'_b = [u_b \ v_b \ 1]^\top$ takes the form:

$$[u_a u_b \ \ u_a v_b \ \ u_a \ \ v_a u_b \ \ v_a v_b \ \ v_a \ \ u_b \ \ v_b \ \ 1] \, \mathsf{vec}(F) = 0.$$

Since in a typical image $u, v \approx 10^2$, the entries of A take values in the interval $[1, \approx 10^4]$, and that of $A^\top A$ in the interval $[1, \approx 10^8]$, resulting in numerical problems. Hartley (1995) suggests the following *normalized eight-point algorithm*:

Algorithm Normalized eight-point algorithm

Input: A set of at least eight corresponding points $\{(m_a'^{(i)}, m_b'^{(i)})\}$ in two images.
Output: The fundamental matrix F.

1: Center the image data at the origin, and rescale it to get a mean distance from the origin of $\sqrt{2}$;
2: use the eight-point algorithm to compute F from the normalized points;
3: enforce the rank-two constraint via SVD;
4: transform the fundamental matrix back to original units: if T_a and T_b are the normalizing transformations for the two images, then the fundamental matrix in original units is $T_b^\top F T_a$.

References

Faugeras, O. D. and Maybank, S. (1990). Motion from point matches: multiplicity of solutions. *International Journal of Computer Vision*, 4(3):225–246.

Fusiello, A. (2018). *Visione computazionale: Tecniche di ricostruzione tridimensionale*. FrancoAngeli.

Fusiello, A., Trucco, E., and Verri, A. (2000). A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*, 12(1):16–22.

Hartley, R. I. (1995). In defence of the 8-point algorithm. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1064–1070.

Huang, T. S. and Faugeras, O. D. (1989). Some properties of the E matrix in two-view motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(12):1310–1312.

Li, H. and Hartley, R. (2006). Five-point motion estimation made easy. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 1, pages 630–633. IEEE.

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293(5828):133.

554SM –Fall 2018

Lecture 6
Stereopsis

END