# **Epipolar Geometry Comparison of SAR and Optical Camera**

Dong Li, Yunhua Zhang

Key Lab of Microwave Remote Sensing (MiRS), Center for Space Science and Applied Research Chinese Academy of Sciences Beijing, China Email: {lidong, zhangyunhua}@mirslab.cn

*Abstract*—In computer vision, optical camera is often used as the eyes of computer. If we replace camera with synthetic aperture radar (SAR), we will then enter a microwave vision of the world. This paper gives a comparison of SAR imaging and camera imaging from the viewpoint of epipolar geometry. The imaging model and epipolar geometry of the two sensors are analyzed in detail. Their difference is illustrated, and their unification is particularly demonstrated. We hope these may benefit researchers in field of computer vision or SAR image processing to construct a computer SAR vision, which is dedicated to compensate and improve our human vision by electromagnetically perceiving and understanding the images.

*Keywords*—Computer vision, computer SAR vision, epipolar geometry, imaging model, optical camera, synthetic aperture radar (SAR), 3D reconstruction.

## **1** Introduction

Human vision is a very intelligent system composed by eyes and brain. The eyes capture the image information, which is then submitted to brain for analysis, learning, recognition, classification, reconstruction, and determination. This system enables us to dynamically interact with the outside world, and it is so natural to us that we often neglect its operation. Computer vision is aimed at duplicating the ability of human vision by electronically perceiving and understanding the images [1]. The optical cameras are often used as eyes to capture images, which are then processed by computer for recognition, analysis, and reconstruction of the objects. This field has received intensive attentions and achieved great development in the past decades, and it has benefited our real life. Nevertheless, it still has a long way to go in order to pursue the intelligence of human vision.

If we replace optical camera with the synthetic aperture radar (SAR), the vision of computer will become completely different. SAR acquires the image of object by actively transmitting an electromagnetic wave with certain frequency and polarization. Such an active operating mode makes it independent of solar illumination and thus allows an all-day imaging. SAR operates in the microwave region of the electromagnetic spectrum (usually between P-band and Ka-band), which can avoid the effects of clouds, fog, rain, and smokes, thus allows an almost all-weather continuous monitoring. The wave-object interaction excites a scattered wave which carries the characteristic information of the object, like the reflectivity, shape, and orientation. By processing the scattering to synthesize a 2D high spatial resolution image, we can achieve a perception about the object. Imaging SAR systems are usually mounted on moving platforms such as airplanes or satellites, and operate in a side-looking geometry. Airborne/Spaceborne SARs provide us microwave visions of the world from aerospace. Using SAR as eves, the vision of computer may be compensated and improved. With the launch and operation of many spaceborne and airborne SAR systems recently, the available high resolution SAR dataset increase dramatically, which makes the joint processing of multiple-view SAR images for accurate understanding and apperception of objects possible. The foundation of computer SAR vision requires us construct the SAR imaging model and epipolar geometry first. The related models for camera are not useful anymore because SAR takes a slant range imaging geometry. In view of this, a concise imaging model and a rigorous epipolar geometry model of SAR were developed recently [2]. This paper is dedicated to give further comparison of SAR imaging and camera imaging from the geometrical viewpoint. Section 2 first concisely presents the imaging model and epipolar geometry of camera. As comparison, the corresponding SAR models are then introduced in Section 3. The difference is revealed and the con-

### 2 Camera Model and Epipolar Geometry

sistency is indicated. Section 4 finally concludes the paper.

#### 2.1 Imaging Model of Camera

Camera acts as a mapping from 3D space to 2D image, which involves in four different coordinate systems, i.e. the world coordinate system  $O_w - X_w Y_w Z_w$ , the camera coordinate system  $O_c - X_c Y_c Z_c$ , the physical image coordinate  $O_p - xy$ , and the digital image coordinate system  $O_r - uv$ . The transformation from  $O_w - X_w Y_w Z_w$  to  $O_c - X_c Y_c Z_c$  usually involves a 3D rotation (*R*) and a translation (*t*) because of the geometrical misalignment between the two systems:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}.$$
(1)

The pinhole camera model then determines the transformation from  $O_c$ - $X_cY_cZ_c$  to  $O_p$ -xy:

$$Z_{c}\begin{bmatrix} x\\ y\\ 1 \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0\\ 0 & f & 0 & 0\\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{vmatrix} X_{c}\\ Y_{c}\\ Z_{c}\\ 1 \end{vmatrix}$$
(2)

where f is the focal length. The finial digital image is a sampling of the physical image, and this can be formulated as:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & -f_u \cot \theta & u_0 \\ 0 & f_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$
(3)

where  $f_u$  and  $f_v$  are the scales in horizontal (u) and vertical (v) image directions, respectively, which relate to the resolutions of the image, and  $\theta$  is the angle between u- and v- axes, which accounts for the fact that the pixel grid may not be exactly orthogonal, and it is usually very close to  $\pi/2$  [3].

Combining (1), (2), and (3), the relation between  $O_w - X_w Y_w Z_w$ and  $O_p - xy$  is obtained:

$$Z_{c}\begin{bmatrix} u\\v\\1\end{bmatrix} = K\begin{bmatrix} R & t\\0^{\mathrm{T}} & 1\end{bmatrix} \begin{vmatrix} X_{w}\\Y_{w}\\Z_{w}\\1\end{vmatrix}$$
(4)

where K is the intrinsic matrix accounting for camera sampling and optical characteristics:

$$\mathbf{K} = \begin{bmatrix} ff_u & -ff_u \cot \theta & u_0 \\ 0 & ff_v / \sin \theta & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$
(5)

Eq. (4) is usually simply expressed as

$$\kappa \tilde{m} = P \tilde{M} = K [R | t] \tilde{M}$$
(6)

where

$$\tilde{\boldsymbol{m}} = \begin{bmatrix} \boldsymbol{u} \\ \boldsymbol{v} \\ 1 \end{bmatrix}, \quad \tilde{\boldsymbol{M}} = \begin{bmatrix} \boldsymbol{X}_{w} \\ \boldsymbol{Y}_{w} \\ \boldsymbol{Z}_{w} \\ 1 \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{R} \mid \boldsymbol{t} \end{bmatrix} = \begin{bmatrix} \boldsymbol{R} & \boldsymbol{t} \\ \boldsymbol{0}^{\mathrm{T}} & 1 \end{bmatrix}. \quad (7)$$

Eq. (6) is just the projective camera model, in which P is the camera matrix, and  $\kappa$  is an arbitrary constant.

#### 2.2 Epipolar Geometry of Camera

Consider the stereo system composed by two cameras shown in Fig. 1, where  $C_1$  and  $C_2$  are the optical centers of the cameras,  $I_1 \leftrightarrow I_2$  is the corresponded projective image pair. Given a pixel  $m_1$  in  $I_1$ , it corresponds to a series of points  $M_1, M_2, \ldots$ in 3D space, and these points lie on the line through  $C_1$  and  $m_1$ . When these points are viewed by the second camera from a distinct position, they will be mapped to the line  $L_2$  in  $I_2$ , and this line is the epipolar line of  $m_1$ . Given a pixel  $m_2$  on  $L_2$ , there then always exists a constraint between  $m_1$  and  $m_2$ . This relation is termed as the epipolar geometry of the stereoscope, also known as the image geometrical warp function because it maps a pixel position in  $I_1$  into a different pixel position in  $I_2$ and forms the so-called image geometrical warp. The epipolar geometry of camera has been extensively studied in the field of computer vision, where the fundamental matrix and homography are the widely-used descriptions of epipolar geometry



Fig. 1. Epipolar geometry of stereo camera vision system.



Fig. 2. Geometrical description of SAR imaging.

when pinhole model is considered. Let the displacement from the first camera to the second be  $(\mathbf{R}, t)$ . Let  $\mathbf{M}$  be the 3D point corresponded to pixels  $m_1$  and  $m_2$ . Without loss of generality, we assume that  $\mathbf{M}$  is expressed in the coordinate system of the first camera. We then have:

$$\begin{cases} \kappa_1 \tilde{m}_1 = K_1 [I \mid 0] \tilde{M} \\ \kappa_2 \tilde{m}_2 = K_2 [R \mid t] \tilde{M} \end{cases}$$
(8)

Eliminating M and the constants  $\kappa_1$  and  $\kappa_2$ , we obtain the following fundamental equation

$$\tilde{\boldsymbol{m}}_{\boldsymbol{2}}^{\mathrm{T}}\boldsymbol{K}_{\boldsymbol{2}}^{-\mathrm{T}}\boldsymbol{T}\boldsymbol{R}\boldsymbol{K}_{\boldsymbol{1}}^{-\mathrm{T}}\tilde{\boldsymbol{m}}_{\boldsymbol{1}}=0$$
(9)

where *T* is an antisymmetric matrix defined by *t*. The fundamental matrix *F* is defined as:

$$F = K_2^{-T} T R K_1^{-T} = K_2^{-T} E K_1^{-T}, \ E = T R$$
(10)

where E is the essential matrix accounting for camera position and orientation in the world coordinate system. Based on the correspondence set  $\{m_1^i \leftrightarrow m_2^i\}$  extracted by certain feature extractor, the fundamental matrix F can be robustly estimated, then the camera matrix P can be simply retrieved from F. And using the method such as triangulation, we can finally locate the 3D point  $M^i$  corresponded to  $m_1^i \leftrightarrow m_2^i$  [3]. This is the task of 3D reconstruction from multiple images.

If point *M* is in a 2D planar scene  $\pi = (n^T, d)^T$ , then  $m_1$  in  $I_1$  is uniquely mapped to  $m_2$  in  $I_2$ . Such mapping can be expressed by the following transformation:

$$\tilde{m}_2 = H\tilde{m}_1 \tag{11}$$

where H is the plane-induced homography, which can be easily obtained from (8) that

$$\boldsymbol{H} = \boldsymbol{K_2} \left( \boldsymbol{R} - \boldsymbol{t} \boldsymbol{n}^{\mathrm{T}} / \boldsymbol{d} \right) \boldsymbol{K_1}^{-1}.$$
 (12)

By fitting  $\{m_1^i \leftrightarrow m_2^i\}$  to (11), an estimation of homography *H* is also achieved, based on which we can then geometrically



Fig. 3. General SAR slant imaging geometry.

align the image pair  $I_1 \leftrightarrow I_2$ . This is the task of image registration, which is the foundation of many applications, such as object tracking and recognition, camera calibration and image reconstruction, as well as the digital elevation model inversion and deformation mapping of the earth surface.

## **3** SAR Imaging Model and Epipolar Geometry

The formulation above indicates that the modeling of epipolar geometry involves in the imaging model of the sensor. SAR acquires image from the slant range, thus the 3D points which are imaged to the same pixel  $m_1$  in  $I_1$  locate in a Doppler circle formed by the intersection of the range sphere and Doppler cone, as shown in Fig. 2. These points are then imaged to a series of pixels in  $I_2$  through the slant projection of the second SAR sensor and compose the epipolar line of  $m_1$ , which is not a simple straight line like  $L_1$  in Fig. 1 any longer. Hence, the description of SAR epipolar geometry with the fundamental matrix is inappropriate because it is only fit for the central projection. SAR image often covers a large ground scene with varied topography, thus the approximation of the 3D ground scene as a planar surface is inaccurate, thus the plane-induced homography is also inappropriate. In order to construct a rigorous description of SAR epipolar geometry, we seek to another idea to achieve it directly from the SAR imaging model. The imaging model maps a 3D point *M* to its projective pixel, so we can relate the two imaged pixels  $m_1$  and  $m_2$  of M under the two SAR projections by combining the two SAR imaging models. And the imaging model should be accurate and concise so as to achieve a rigorous and analytical epipolar modeling. The existing SAR imaging models can be generally attributed into two categories, i.e. the physical model and empirical model. The physical model takes into account of several aspects that influence the acquisition procedure based on the range-Doppler equations (RDEs):

$$\begin{cases} R = \sqrt{(X_S - X_P)^2 + (Y_S - Y_P)^2 + (Z_S - Z_P)^2} \\ \frac{V_X (X_S - X_P) + V_Y (Y_S - Y_P) + V_Z (Z_S - Z_P)}{R} = -\frac{f_d \lambda}{2} \end{cases}$$
(13)

where *R* is distance from object  $(X_P, Y_P, Z_P)$  to antenna phase center (APC)  $(X_S, Y_S, Z_S)$ ,  $(V_X, V_Y, V_Z)$  is the velocity of platform,  $\lambda$  is the wavelength of the transmitted wave, and  $f_d$  is the Doppler frequency. Theoretically, we can obtain an accurate SAR epipolar model from RDEs of the two SARs. But the model may not be concise because of the complex nonlinearity of RDEs, which will impact the further application. The empirical model is used when the system parameter, imaging geometry, and physical model are unavailable, and the polynomial and rational polynomial are often used:

$$P = \sum_{i=0}^{m} \sum_{j=0}^{n} \sum_{k=0}^{p} a_{ijk} X^{i} Y^{j} Z^{k}, \quad R = \frac{\sum_{i=0}^{m} \sum_{j=0}^{n} \sum_{k=0}^{p} a_{ijk} X^{i} Y^{j} Z^{k}}{\sum_{i=0}^{m} \sum_{j=0}^{n} \sum_{k=0}^{p} b_{ijk} X^{i} Y^{j} Z^{k}}$$
(14)

which are in fact the extensions of homography and collinearity equations obtained from central projection, respectively:

$$\begin{cases} H = a \left( X - X_0 \right) + b \left( Y - Y_0 \right) + c \left( Z - Z_0 \right) + d \\ C = \left( -f \right) \frac{m_{11} \left( X - X_0 \right) + m_{12} \left( Y - Y_0 \right) + m_{13} \left( Z - Z_0 \right)}{m_{31} \left( X - X_0 \right) + m_{32} \left( Y - Y_0 \right) + m_{33} \left( Z - Z_0 \right)} . \end{cases}$$
(15)

However, different from the central projection of camera, the equivalent projection center for slant range SAR imaging is not fixed, i.e. SAR is a variable focus system or multi-central projection system, thus the empirical model is also inaccurate.

#### 3.1 Concise Imaging Model of SAR

Here we consider the general SAR imaging geometry shown in Fig. 3. The rigorous modeling of SAR epipolar geometry requires the imaging model be concise and accurate. In order to achieve this, we also construct four coordinate systems, i.e. the global coordinate system **O-XYZ**, the platform coordinate system **O'-X'Y'Z'**, the imaging coordinate system **o-xyz**, and the image coordinate system o<sub>i</sub>-uv. O-XYZ is similar to the world coordinate system  $O_w$ - $X_w Y_w Z_w$ . Assume radar moves along a straight track of height H paralleling to the ground plane XOY, which indicates the influence from the curvature of earth and track is neglected, thus we mainly focus on the airborne SAR system. Nevertheless, it may also hold for the spaceborne SAR system if we can compensate those nonideal influences by using high precise platform-borne GPS and INS beforehand. O'-X'Y'Z' is used to characterize the attitude of platform (its counterpart is the camera coordinate system  $O_c$ - $X_c Y_c Z_c$ ), where **O'** is located at  $(T_X, T_Y, H)$  in **O-XYZ** representing the initial APC, X' denotes the flight direction of the platform, Z' is parallel to Z, and Y' is orthogonal to X' and Z'. If the flight direction X' is  $\beta$  deviated from X, the transformation between *O'-X'Y'Z'* and *O-XYZ* can thus be expressed as:

$$\begin{pmatrix} X' \\ Y' \\ Z' \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{\boldsymbol{\beta}} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix} \begin{pmatrix} X - T_{X} \\ Y - T_{Y} \\ Z - H \end{pmatrix} \text{ with } \mathbf{R}_{\boldsymbol{\beta}} = \begin{pmatrix} \cos \beta & \sin \beta \\ -\sin \beta & \cos \beta \end{pmatrix}. (16)$$

We further consider an imaging geometry that the antenna has a squint angle of  $\alpha$  which anti-clockwisely rotates Y' to the incidence plane. This is a special characteristic of SAR imaging. For squint SAR, we should compensate the scattering to the zero Doppler centroid first. For the sake of convenience, the imaging coordinate system *o-xyz* is further defined, where *o* is located at ( $T_X$ ,  $T_Y$ , 0) in *O-XYZ*, *z* is parallel to *Z*, *y* is parallel to the ground projection of the antenna boresight, and *x* is orthogonal to *y* and *z*. The compensation results in an anticlockwise rotation of  $\alpha$  from *X'* to *x*, thus the relation between *o-xyz* and *O'-X'Y'Z'* is:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{R}_{\alpha} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix} \begin{pmatrix} X' \\ Y' \\ Z' + H \end{pmatrix} \text{ with } \mathbf{R}_{\alpha} = \begin{pmatrix} \cos \alpha & \sin \alpha \\ -\sin \alpha & \cos \alpha \end{pmatrix}.$$
(17)

By combining (16) and (17), we can obtain the transformation from *O-XYZ* to *o-xyz* 

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{pmatrix} \begin{pmatrix} X - T_{X} \\ Y - T_{Y} \\ Z \end{pmatrix} \text{ with } \mathbf{R} = \begin{pmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{pmatrix}$$
(18)

where  $\varphi (= \alpha + \beta)$  denotes the anti-clockwise rotation from *X* to *x*. After these, SAR imaging can be modeled as a geometrical projection from ground plane to slant plane. Let *C* be a 3D point within the radar beam with coordinates of (*X*, *Y*, *Z*) and (*x*, *y*, *z*) in *O-XYZ* and *o-xyz*, respectively. After slant projection, *C* is mapped to *C'* (*x<sub>p</sub>*, *y<sub>p</sub>*, *z<sub>p</sub>*). From the projection geometry in Fig. 3, we can easily obtain that

$$x_p = x, y_p = P'C' = P'C = y\sin^{-1}\theta, z_p = H.$$
 (19)

where  $\theta$  is the local radar incidence related to the position and height of *C*. Thus the relation between slant projective plane and ground plane can be written as:

$$\begin{pmatrix} x_p \\ y_p \end{pmatrix} = \boldsymbol{M} \cdot \begin{pmatrix} x \\ y \end{pmatrix} \text{ with } \boldsymbol{M} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^{-1} \theta \end{pmatrix}.$$
(20)

The counterpart of this transformation in camera model is (2). The final SAR image is the sampling of the projective plane. The image is defied in the image coordinate system  $o_r uv$ , the transformation between pixel (u, v) and projection  $(x_p, y_p, z_p)$  can then be expressed as:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} s_x & 0 \\ 0 & s_y \end{pmatrix} \begin{pmatrix} x_p - t_x \\ y_p - t_y \end{pmatrix} = \mathbf{S} \cdot \begin{pmatrix} x_p - t_x \\ y_p - t_y \end{pmatrix}$$
(21)

where  $(t_x, t_y, H)$  is the location of  $o_i$  in *o-xyz*,  $s_x$  and  $s_y$  are the scales related to the azimuth and range pixel sizes  $\Delta_a$  and  $\Delta_r$ , respectively

$$S_x = 1/\Delta_a = 2s_a/L_e, \ s_y = 1/\Delta_r = 2s_r B/c$$
 (22)

where  $L_e$  is the effective antenna aperture, *B* is the bandwidth of the transmitted signal,  $s_a$  is the azimuth oversampling rate related to the PRF of the system,  $s_r$  is the range oversampling rate, and *c* is the velocity of light. The transformation in (21) is similar to (3) of camera model. However, the pixel grid in SAR image can be generally kept exactly orthogonal, thus the retinal distortion is neglected here.

By combining (18), (20), and (21), we have

$$\begin{pmatrix} u \\ v \end{pmatrix} = \mathbf{S} \cdot \mathbf{M} \cdot \mathbf{R} \cdot \begin{pmatrix} X - T_X \\ Y - T_Y \end{pmatrix} - \begin{pmatrix} s_x t_x \\ s_y t_y \end{pmatrix}.$$
 (23)

Equation (23) finally relates a 3D point to its projective pixel in SAR image. Under the assumption that radar moves along a track paralleling to the ground plane, the RDEs is in fact consistent with (23) because any approximation has not been used when deriving the relation. Nevertheless, we decompose the complex RDEs into the multiplication of three simple matrices of physical significance based on the transformations among four different coordinate systems, which helps us perform a concise and accurate result. The model involves in the system parameters  $s_x$  and  $s_y$ , the imaging geometry parameters  $\varphi$ ,  $T_x$ ,  $T_y$ ,  $t_x$ , and  $t_y$ , as well as the object parameter  $\theta$ . It is interesting to observe that (23) is similar to the linear camera model of (6): here **R** as well as  $T_X$ ,  $T_Y$ ,  $s_x t_x$ , and  $s_y t_y$  correspond to (1) denoting the transformation from world coordinate system to camera coordinate system by rotation and translation, *M* corresponds to (2) which indicates the transformation from camera system to physical image coordinate system by pinhole model, and S corresponds to (3) denoting the transformation from image physical system to pixel coordinate system by digital sampling. Therefore, the obtained model enables us to geometrically unify SAR imaging and camera imaging. Besides this, from (23) one can see that the point position (X, X)Y) is explicitly related to the pixel position (u, v), but the point elevation Z is implicit in the local incidence  $\theta$ , thus the model may also enable a flexible strategy to model the epipolar geometry and to reconstruct the object.

#### 3.2 Rigorous Epipolar Geometry of SAR

The side-looking of SAR makes the epipolar geometry depiction in terms of fundamental matrix and homography inappropriate, we thus turn to construct the SAR epipolar geometry directly from the imaging model. This kind of epipolar geometry description is less used for camera because the fundamental matrix and homography are both good enough, but it facilitates us to model the rigorous epipolar geometry for SAR from the developed concise imaging.

We consider a general stereoscopic configuration here. Let  $I_1$ and  $I_2$  be an image pair acquired by different SAR systems from different imaging geometries. For a 3D point (*X*, *Y*, *Z*) in *O*-*XYZ*, if its two projective pixel positions in  $I_1$  and  $I_2$  are ( $u_1$ ,  $v_1$ ) and ( $u_2$ ,  $v_2$ ), respectively, according to (23) we obtain

$$\binom{X}{Y} = \mathbf{R}_{\mathbf{1}}^{-1} \mathbf{M}_{\mathbf{1}}^{-1} \mathbf{S}_{\mathbf{1}}^{-1} \binom{u_1}{v_1} + \mathbf{R}_{\mathbf{1}}^{-1} \mathbf{M}_{\mathbf{1}}^{-1} \mathbf{S}_{\mathbf{1}}^{-1} \binom{s_{x1} t_{x1}}{s_{y1} t_{y1}} + \binom{T_{X1}}{T_{Y1}}$$
(24)

where the subscript 1 indicates the parameters of  $I_1$ . Based on (23) and (24), for pixel  $(u_2, v_2)$  of  $I_2$  we can have

$$\begin{pmatrix} u_{2} \\ v_{2} \end{pmatrix} = S_{2}M_{2}R_{2}R_{1}^{-1}M_{1}^{-1}S_{1}^{-1}\begin{pmatrix} u_{1}+s_{x1}t_{x1} \\ v_{1}+s_{y1}t_{y1} \end{pmatrix}$$

$$+ S_{2}M_{2}R_{2}\begin{pmatrix} T_{X1}-T_{X2} \\ T_{Y1}-T_{Y2} \end{pmatrix} - \begin{pmatrix} s_{x2}t_{x2} \\ s_{y2}t_{y2} \end{pmatrix}$$

$$(25)$$

here the subscript 2 indexes the parameters of  $I_2$ . By eliminating the object position (X, Y), we can then relate the two pixels. Equation (25) can be further rearranged as (26), shown at the top of the next page, where  $A = S_2 M_2 R_2 R_1^{-1} M_1^{-1} S_1^{-1}$ ,  $\Delta \varphi$  (=  $\varphi_2 - \varphi_1$ ) is the rotation between the two imaging systems,  $B_x$ and  $B_y$  are the projections of  $B_X$  (=  $T_{X1} - T_{X2}$ , denotes the initial separation between the two imaging systems in X-direction)

$$\begin{pmatrix} u_{2} \\ v_{2} \end{pmatrix} = \mathbf{A} \begin{pmatrix} u_{1} \\ v_{1} \end{pmatrix} + \begin{pmatrix} t_{u} \\ t_{v} \end{pmatrix} = \begin{pmatrix} s_{x1}^{-1} s_{x2} \cos \Delta \varphi & s_{y1}^{-1} s_{x2} \sin \theta_{1} \sin \Delta \varphi \\ -s_{x1}^{-1} s_{y2} \sin^{-1} \theta_{2} \sin \Delta \varphi & s_{y1}^{-1} s_{y2} \sin \theta_{1} \sin^{-1} \theta_{2} \cos \Delta \varphi \end{pmatrix} \begin{pmatrix} u_{1} \\ v_{1} \end{pmatrix} + \begin{pmatrix} s_{x2} \left( t_{x1} \cos \Delta \varphi + t_{y1} \sin \theta_{1} \sin \Delta \varphi + B_{x} - t_{x2} \right) \\ s_{y2} \sin^{-1} \theta_{2} \left( -t_{x1} \sin \Delta \varphi + t_{y1} \sin \theta_{1} \cos \Delta \varphi + B_{y} - t_{y2} \sin \theta_{2} \right) \end{pmatrix}.$$

$$(26)$$

and  $B_Y (= T_{Y1} - T_{Y2})$ , denotes the initial separation between the two imaging systems in *Y*-direction) in *x*- and *y*-directions of the second SAR imaging system, and they denote the alongtrack and cross-track baselines, respectively. Equation (26) is an affine transformation which models the epipolar geometry of a general SAR stereo. Interestingly, it is similar to the plane-induced homography in (12): here  $S_1M_1$  and  $S_2M_2$  correspond to the intrinsic matrices  $K_1$  and  $K_2$  of the two cameras,  $R_2 R_1^{-1}$  as well as  $t_u$  and  $t_v$  correspond to the rotation and translation  $\mathbf{R} - t\mathbf{n}^{T}/d$  between the two cameras. However, different from the fundamental matrix and homography which are independent of object, the SAR epipolar geometry in (26) is object-dependent because the local radar incidences  $\theta_1$  and  $\theta_2$ vary with each 3D point. Nevertheless, (20) and (23) show that, for slant range imaging, the position and elevation of a 3D point are wrapped into the imaged pixel and related to the local incidence. Therefore, given the imaging parameters and pixel correspondences, we can obtain a retrieval of incidence from (26), then the 3D geometry of object may be achieved based on the projective pixel positions. Therefore, the image reconstruction in computer SAR vision seems more straightforward than that in computer optical vision, as detailed in [2].

## 4 Conclusions

The slant range imaging of SAR makes the equivalent projection center unfixed. SAR is therefore a variable focus system or multi-central projection system. This paper is dedicated to give a comparison of SAR imaging and camera imaging from the geometrical point of view. A unified expression of camera imaging model and SAR imaging model is obtained. Nevertheless, the side-looking makes the SAR imaging model vary with each object position, besides introducing an extra squintrelated rotation. We thus cannot use a fixed model to express the mapping from 3D space to 2D SAR image. The epipolar geometry models the relation between the two pixels of a 3D point projected by a stereoscope. The central projection of camera enables a pixel position in the first image to be corresponded to a straight line in the second image. Such relation is described in terms of the fundamental matrix. It can be also expressed by the homography if the considered scene is planar. However, these two descriptions of epipolar geometry are all inappropriate for SAR because slant range imaging makes the epipolar line not be straight anymore, and SAR image often covers a large ground scene with varied topography which cannot be approximated as a planar surface. Hence, we turn to construct the rigorous SAR epipolar geometry directly from the imaging model. Nevertheless, its unification with the plane-induced camera homography is also clear. The obtained SAR epipolar geometry also varies with each object position, i.e. we cannot use a fixed homography to model the geometric wrapping of two SAR images. This makes the image registration more difficult. However, such object-dependent epipolar geometry is welcome for the retrieval of 3D geometry of object. Hence, the 3D reconstruction of object in computer SAR vision thus seems more straightforward than that in computer optical vision.

The consistency on imaging model as well as epipolar geometry indicates the geometrical unification of SAR imaging and camera imaging in a sense. Nevertheless, this does not mean that we can unify the two sensors in physics. Our focus in this paper is paid on the geometrical part of the imaging. In fact, the side-looking also impacts the physical appearance of SAR image. The detailed comparison of SAR imaging and camera imaging from the electromagnetic scattering and signal point of view was presented in [4].We hope such geometrical consistency will benefit researchers in field of computer optical vision or SAR image processing to construct a computer SAR vision. As an interdiscipline of the two fields, computer SAR vision is dedicated to improve and compensate human vision by electromagnetically perceiving and understanding the images, which will make our view of the world more colorful.

## Acknowledgements

This work was supported by the Youth Innovation Promotion Association, Chinese Academy of Sciences under the grant no. Y42133A39S.

## References

- M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 3rd ed. Toronto, Canada: Thomson, 2008.
- [2] D. Li and Y. Zhang, "A rigorous SAR epipolar geometry modeling and application to 3D target reconstruction," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 6, no. 5, pp. 2316-2323, Oct. 2013.
- [3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2003.
- [4] Y. Zhang and J. Jiang, "Why optical images are easier to understand than radar images? –From the electromagnetic scattering and single point of view," in *Proc. PIERS*, Guangzhou, China, 2014, pp. 1411-1415.