

## 合成孔径雷达三维成像中的视觉语义浅析

胡占义\*

(中国科学院自动化研究所模式识别国家重点实验室 北京 100190)

**摘要:**“合成孔径雷达微波视觉三维成像”，从概念上说，旨在将“视觉语义”引入到合成孔径雷达的成像模型中，以期提高三维成像的质量。对层析合成孔径雷达(TomoSAR)来说，“视觉语义”的引入可望有效减少TomoSAR所需的观测次数。然而，什么是“视觉语义”？从视觉感知的途径看，“单眼”和“双眼”均可以从场景感知三维结构信息；从场景内容看，不同的人对同一幅图像会有不同感受；从视觉神经加工机理看，三维信息加工和二维信息加工也存在一些本质差异。另外，人类视觉感知普遍存在错觉(illusion)现象。那么，到底什么类型的“视觉语义信息”可望在计算的层次上有助于微波三维成像呢？如何借鉴计算机视觉的理论和方法来提取微波三维成像中有用的“视觉语义”信息呢？该文对这些问题进行了一些初步探讨。

**关键词:** 视觉通道；微波视觉；层析SAR

中图分类号: TP391

文献标识码: A

文章编号: 2095-283X(2022)01-0020-07

DOI: 10.12000/JR21149

**引用格式:** 胡占义. 合成孔径雷达三维成像中的视觉语义浅析[J]. 雷达学报, 2022, 11(1): 20–26. doi: 10.12000/JR21149.

**Reference format:** HU Zhanyi. A note on visual semantics in SAR 3D imaging[J]. *Journal of Radars*, 2022, 11(1): 20–26. doi: 10.12000/JR21149.

## A Note on Visual Semantics in SAR 3D Imaging

HU Zhanyi\*

(National Laboratory of Pattern Recognition, Institute of Automation,  
Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** Conceptually speaking, Synthetic Aperture Radar (SAR) microwave vision 3D imaging refers to fusing visual semantics into the SAR 3D imaging process to enhance the 3D imaging quality. For SAR Tomography (TomoSAR), it specifically means to reduce the needed observations by fully exploiting SAR visual semantics. However, what does it mean by visual semantics? From the viewpoint of visual perception, 3D structural information could be perceived from either monocular image or binocular images and the same scene could be perceived differently by different people. From the viewpoint of neurophysiology, depth perception from binocular or monocular vision has fundamentally different mechanism. Besides visual illusion phenomenon is omnipresent in daily life. Hence what kinds of visual semantics could be helpful for SAR 3D imaging from the computational point of view? What could be learnt from computer vision community to extract useful visual semantics from SAR images? This short note presents some preliminary discussions on such issues, a purely personal view on such vast topics.

**Key words:** Visual pathway; Microwave vision; TomoSAR

### 1 背景与引言

2019年，国家自然科学基金委经过反复论证，

收稿日期: 2021-10-09; 改回日期: 2021-11-20; 网络出版: 2021-12-10

\*通信作者: 胡占义 huzy@nlpr.ia.ac.cn

\*Corresponding Author: HU Zhanyi, huzy@nlpr.ia.ac.cn

基金项目: 国家自然科学基金(61991423, 61772444)

Foundation Items: The National Natural Science Foundation of China (61991423, 61772444)

责任编辑: 丁赤旸 Corresponding Editor: DING Chibiao

将“合成孔径雷达微波视觉三维成像理论与应用基础”设立为重大研究项目内容。该项目的重大原始创新是将“微波视觉”概念引入到SAR三维成像框架中，以期在减少SAR观察次数的情况下，仍能有效处理SAR图像的叠掩现象和散射体的高程信息复原能力。笔者觉得，“微波视觉”目前仍是一个框架性概念，“微波视觉语义”也是一个“内涵”比较广的概念，需要对这些内容进行探讨和具体化。笔者长期从事计算机视觉研究，对生物视觉

也有一些了解,但对SAR成像和处理了解非常浅薄。借此专刊提供契机,谈一点自己对这些问题的粗浅看法,不妥之处敬请批评指正。另外,本文主要介绍一些笔者的看法,而不是系统介绍相关内容,所以尽量减少了对相关文献的索引。

## 2 什么是视觉和视觉语义

在探讨“微波视觉”之前,先讨论一下什么是“视觉”。关于什么是“视觉”这个问题,从古希腊亚里士多德起,人们的讨论就没有停止过。笔者觉得,“计算视觉”(computational vision)的奠基人David Marr<sup>[1]</sup>(马尔)在其*vision*一书中,给出的定义可能是对“视觉”最精炼和简洁的描述。马尔认为,“视觉”就是通过看来确定什么东西在什么地方(Vision is to know what is where by looking)。尽管很多人觉得“视觉”远不止马尔描述的“what”和“where”问题,但“什么东西”在“什么地方”至少是视觉的基本功能。把视觉功能过分扩大就会与脑功能混淆,如盲人都具备的能力,似乎不太合适。笔者觉得,“视觉”应该首先研究“大脑视觉皮层”的功能,而不宜重点研究涉及多通道融合的脑皮层区域的功能,否则,“视觉”与“脑科学”就没有多少区别了。

### 2.1 物体视觉和空间视觉

为了适应白天和黑夜光强的剧烈变化,人类视觉系统进化成了“日视”和“夜视”两套成像系统。人类约1.2亿个感光细胞(photoreceptors)中,约1.1亿为杆状细胞(rod),600万~700万为锥状细胞(cone)。杆状细胞主要负责夜视,锥状细胞大多可以感知颜色,用于“日视”。视网膜(retina)是成像部位,对图像进过初级加工后,如去噪、对比度增强等,然后将信号传到枕叶(occipital lobe:图1中的绿色区域)的视觉初级加工区域(V1, V2区等)进行加工处理。

如图1所示,信号经过视皮层初级处理(如边缘提取、运动检测、视差估计等)后,主要分成两个加工通道,一个是腹部通道(ventral pathway:绿色到蓝色区域虚线),主要负责物体识别,称为“物体视觉”(object vision)。另一个是背部通道(dorsal pathway:绿色到红色区域虚线),主要负责“操作物体”的视觉,由于操作物体必然涉及空间位置和距离等信息,所以称为“空间视觉”(spatial vision)。

### 2.2 深度感知:单目感知和双目感知

由于本文主要关心“图像三维视觉语义”,下面对单目深度感知和双目深度感知进行一些简单介绍。

目前神经科学对“双目立体视觉”(binocular stereo)的机理相对比较清楚。单眼信号首先在视觉

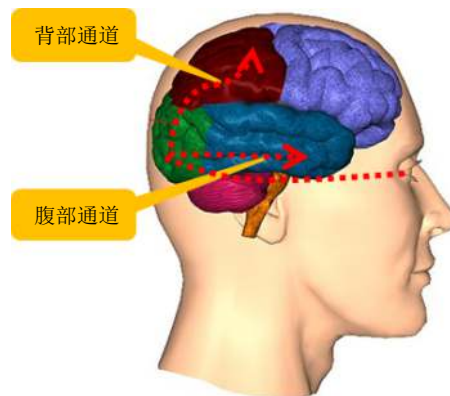


图1 视觉腹部通道和背部通道。腹部通道主要负责物体视觉,背部通道主要负责空间视觉

Fig. 1 Visual ventral pathway and dorsal pathway: Ventral pathway is mainly for object vision, dorsal pathway for spatial vision

V1区进行融合,并对绝对视差(absolute disparity)进行加工,然后在后续皮层进一步对绝对视差精化和相对视差计算。腹部通道和背部通道均涉及视差处理,但到目前为止,人们还没有发现任何一个脑皮层区域“专门用来处理视差”。双目视差处理目前主要有2种计算模型:一种是1990年Ohzawa等人<sup>[2]</sup>提出的视差能量模型(disparity energy model),一种是Haefner和Cumming<sup>[3]</sup>于2008年提出的扩展的视差能量模型:2SU模型。由于人的双眼间距很小,外界环境在双眼视网膜上的成像基本上相差一个很小的平移,所以视差能量模型本质上是多个神经元对“图像相关”计算的一种模型。

很显然,单目也可以进行场景深度感知,仅仅是感知的精度要较双目差一些。目前,就笔者所知,还没有任何关于单目深度感知的相关神经加工机理的报道。目前的一些线索大多是“心理学”的一些实验结论<sup>[4]</sup>。如:相对尺寸“relative size”(同样大小的两个物体,看上去大的物体在前,小的在后),纹理梯度(Texture Gradient)(梯度大的区域在前面),线性透视(Linear Perspective)(平行线的投影越到后面变得越窄)等。这些线索还很难上升成“计算原理”。因为这些都是些“感觉”。三维成像是现实的复原,而三维感觉却可以远离现实。目前的很多“虚拟或增强现实”,事实上都在给予人们对“非现实”的“现实感”,本质上都是一些错觉。

### 2.3 视觉错觉(visual illusion)

人类视觉系统既可以从“无真实三维信息的图像”感觉到三维信息,也可以从“包含真实三维信息的图像”得到错误感觉。如图2的线画图,人们可以产生三维感觉。图3的“The Ames room illu-

sion”，两个人的真实身高差不多，但人们都会有“前面人高，后面人矮”的感觉。

“视觉语义”就是对场景感知信息在“语义层次上”的描述，即在“概念”层次上的一种描述。“错觉”就会导致“错误的视觉语义”。人们对场景有多种感受，但不是所有感受都可以上升到概念层次。另外，感知得到的视觉语义并不见得都是对真实场景的表述。如人们看到的颜色本质上是对波长信息的语义表述。如何在SAR三维成像中利用三维视觉语义信息，以提高三维成像质量，仍需要在框架、理论和算法3个层次上进行深度探讨。

### 3 什么是“微波视觉语义”

什么是“微波视觉语义”？笔者觉得就是人们从微波图像“感知”得到的“场景语义信息”。也就是人们从微波图像“直接看到的”场景语义信息。尽管SAR是距离成像，存在叠掩等光学成像系统不存在的特有现象，但人们直接从SAR图像也确实可以感知到一些场景三维结构信息，如从图4的SAR图像中，人们可以感知到的船和桥的一些三维结构信息。

本文笔者仅仅讨论SAR三维成像中的视觉语义，还不是更广泛意义下的微波视觉语义。由于“视

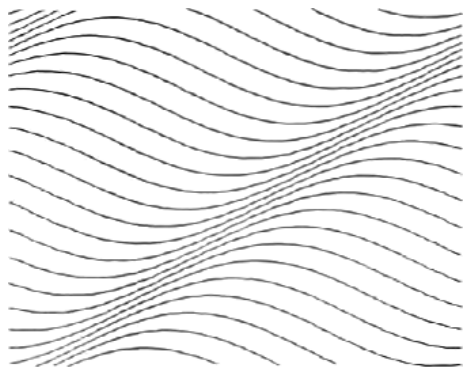


图2 从线画图可以产生三维感觉<sup>[1]</sup>

Fig. 2 Human could perceive 3D shape from line drawing<sup>[1]</sup>



图3 图中两个人身高感觉存在明显差异

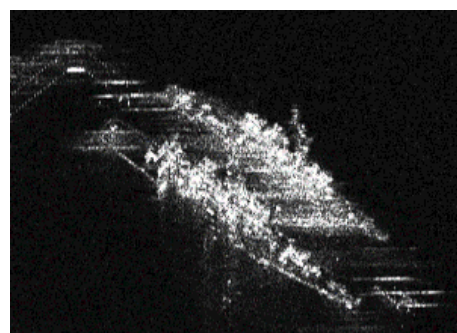
Fig. 3 The Ames room illusion. Two women in the picture have similar heights, but perceived very differently

觉”包含“视觉感知”和“视觉认知”。计算机视觉传统意义下主要研究视觉感知问题，而视觉认知覆盖了更广泛的概念(如回想视觉事件、视觉概念形成、视觉事件推理等)，且笔者觉得，视觉认知问题似乎与其他感觉通道信息的认知机理也没有本质区别(如视觉事件推理与听觉事件推理似乎没有本质区别)。正像计算机视觉领域一样，把“计算机视觉”范畴过度扩展，就会与“图像理解”，“图像分析”，“视频分析”产生混淆，笔者觉得，“微波视觉”似乎也应该避免类似问题，否则也会存在与微波图像理解、微波图像分析混淆的可能。

笔者觉得，“SAR三维成像中的视觉语义”，就是指如何利用从SAR图像中感知的语义信息来“增强SAR图像的三维成像能力”。也就是说，在传统SAR三维成像中，如何通过增加“视觉语义”约束，来提高SAR三维成像的性能。

### 4 如何利用视觉语义提高SAR三维成像：处理框架

正像前面所述，“三维成像”是对真实三维场景的复原，“视觉语义”是对真实场景的“主观感



(a) TerraSAR-X 聚束模式下的0.5 m 图像  
(a) TerraSAR-X 0.5 m image



(b) 中科院空天院的机载0.3 m SAR图像  
(b) 0.3 m image of AIRI aerial system of Chinese Academy of Sciences

图4 从图(a)可以感知到船的一些三维结构信息；图(b)可以感知到桥的一些三维结构

Fig. 4 3D ship structural information could be perceived from (a); Bridge 3D shape could be clearly perceived from (b)

受”，而主观感受可能产生错觉。所以如何在SAR三维成像中融合合适的视觉语义信息，需要探索一套计算框架、计算理论和计算方法。下面主要围绕本重大项目的“层析SAR”(TomoSAR)问题进行一些讨论。

TomoSAR 是一种恢复高程信息的有效技术途径<sup>[5]</sup>(这里的高程信息主要指位置信息)。本项目的-一个主要目标在于“如何融合视觉语义信息，以减少TomoSAR的观测次数，实现SAR快速三维成像”，以有效解决传统TomoSAR周期长、成本高，不利于时效性要求较高的应用等问题。鉴于TomoSAR框架下文献中对单个像元的高程恢复问题已研究了20多年，如以谱分析方法<sup>[6]</sup>和压缩感知方法<sup>[7]</sup>为代表的两大类方法，笔者认为，融合视觉语义的TomoSAR研究，首先应该在处理框架上有别于传统方法。应该从“单像元”处理转变到“图像区域”处理，应该从“前馈式”处理方式转变到“反馈式”处理方式。关于图像区域处理，文献中已有一些报道，如Rambour等人介绍的空间正则途径<sup>[8]</sup>。这里的“前馈式”处理是指从SAR图像一次准确估计高程信息的过程，并不意味着在估计过程中没有迭代计算。“反馈式”处理是指把“初始估计的粗略高程结果”再反馈到下次估计，逐次迭代求精的处理方式。

#### 4.1 基于鲁棒统计的迭代式处理框架

目前文献中的方法，基本上是增加各种约束，如基于压缩感知(Compressed Sensing, CS)的稀疏性约束，将“单个像元内所有散射体的高程信息一次准确恢复”的过程。这是一种典型的“前馈处理”方式。任何图像都不是随机分布的，这种以像元为基本处理单元的方式，既没有考虑像元邻域关系，也没有考虑“场景特有的结构先验知识”。TomoSAR旨在恢复未知的三维场景结构，TomoSAR处理中如何利用“待处理场景特有的结构先验”似乎就成了一个“鸡与蛋”的关系。解决“鸡与蛋”关系，计算中采用的是“迭代”策略，其核心假定是：当没有场景结构先验时，TomoSAR初始恢复的高程不可能非常准确，但存在“一定程度的可靠性”。这些具有一定可靠性的高程信息，特别是一个“区域”对应的粗略高程信息，构成了对场景结构恢复中下一次迭代的“有效先验”。通过将-这些先验融入到下一次TomoSAR的迭代中，可望有效提升TomoSAR的性能。著名的Adaboost分类方法<sup>[9]</sup>是这方面的一个典型代表。每个弱分类器，只要其分类的正确概率大于0.5，多个弱分类器的组合，就可以构成一个性能优良的强分类器。

在这种迭代框架下，随着迭代的进行，场景的高程信息会恢复得越来越准确。这种迭代式估计方法，其合理性支撑理论是鲁棒统计理论，如RANSAC方法<sup>[10]</sup>。图5给出一种TomoSAR迭代估计框架。

目前这种迭代框架下的求解TomoSAR的方法还不多见。Rambour等人<sup>[11]</sup>给出的REDRESS算法，利用城镇场景的特性，通过graph-cut对初始CS框架下估计的高程进一步优化后，进而利用场景信息改变CS中的稀疏性惩罚系数的方法，本质上是一种迭代框架下的TomoSAR方法。

#### 4.2 伪多尺度处理框架

多尺度方法是信息领域一种广泛使用的方法<sup>[12]</sup>，如图像的金字塔表示。多尺度理论在特征提取中的假定：真实的特征在不同尺度下均存在，虚假特征仅仅在某个尺度下出现。另外，多尺度理论也表明，一个特征具有其固有的尺度(intrinsic scale)，也就是说，特征在其固有的尺度下更容易可靠提取，如公路不宜在“厘米分辨率”的图像上提取。TomoSAR也可以在多尺度处理框架下进行处理。如利用谱分析方法在低分辨率下先得到一个粗略高程估计，然后利用该粗略估计的信息作为先验，进一步在压缩感知框架下精化估计结果。这种途径可以将TomoSAR处理的两大途径：谱分析方法和压缩感知方法结合起来，同时可以在“反馈式”处理模式下进一步优化。图6为一种伪多尺度处理框

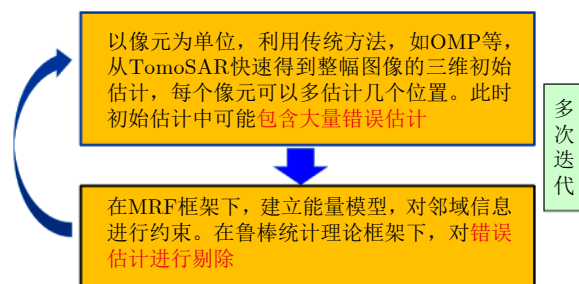


图5 TomoSAR迭代处理框架

Fig. 5 TomoSAR iterative framework

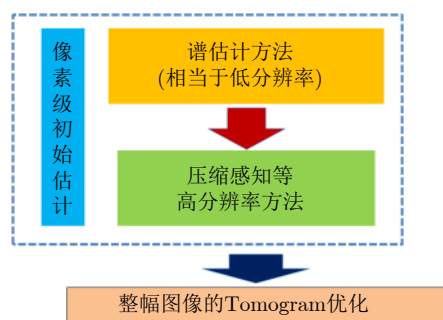


图6 TomoSAR的伪多尺度处理框架

Fig. 6 TomoSAR pseudo-multi-scale framework

架。这里“伪”是为了表明这不是一种真正符合“多尺度理论”的方法。

“迭代框架”、“多尺度框架”一定意味着更长的处理时间吗？计算机视觉领域的大量方法和应用表明<sup>[13,14]</sup>，“多尺度”和“迭代”在估计精度提高的情况下，计算时间反而可以降低。TomoSAR处理与其他图像应用相比，在这方面似乎也不应该存在本质区别。

## 5 如何利用视觉语义提高SAR三维成像：技术途径和算法

TomoSAR融合先验知识，包括视觉语义知识，目前的基本处理途径如式(1)所示，即在TomoSAR传统表述方程中增加视觉语义约束，以提高三维成像的质量。

$$X_M = \operatorname{argmin}_{X_M} \left\{ \underbrace{|Y_N - A_{N \times M} X_M|_{\text{norm}}}_{\text{TomoSAR成像方程}} + \underbrace{\alpha f(X_M)}_{\text{各种约束}} \right\} \quad (1)$$

这种途径在理论上具有融合各种先验知识的潜力和灵活性。式(1)中 $f(X_M)$ 既可以包含连续变量约束，也可以包含离散变量约束，既可以表示确定性正则化(deterministic regularization)，也可以表示统计性正则化(statistical regularization)，同时可以表示更一般的语义正则化(semantic regularization)(如 $X_M$ 位于同一空间水平面上)。既可以表示对单像元的约束，也可以表示空间邻域像元之间的约束。笔者觉得，SAR三维成像中的视觉语义可以在语义正则化框架下进行描述和体现。根据前面对视觉语义的讨论，SAR三维成像中的视觉语义应该是体现场景结构的语义信息，特别是组成场景的几何基元信息，如空间线段、面片等基元，其位置和朝向以及物体类别等信息。另外，这些几何基元信息可以通过“机器学习的途径”来提取。这里需要指出的是，“语义约束”从能量模型的观点看，一般是一个“高阶能量项”(high-order energy model)。从条件随机场(conditional Random Field)能量优化理论知道<sup>[15]</sup>，除了很少的一些高阶能量模型外，一般的含高阶能量项的优化问题都是一个NP-Hard问题。所以，上述框架下设计“约束项”时，一定要考虑对应的求解问题。否则，会出现目前很多“设计了一个复杂优美的能量模型，用简化方法进行了求解，得到了与所设计的能量模型关系不大的结果”的怪现象。

深度学习的进展使得计算机视觉研究发生了“变革性”进展。深度学习可以用在TomoSAR中吗？目前见到的报道并不多，少有的几项工作如

Costante等人<sup>[16]</sup>直接从SAR图像推断DEM的工作，Budillon等人<sup>[17]</sup>直接用深度学习反演TomoSAR，以及Wu等人<sup>[18]</sup>在CS粗估计下进一步利用DNN进行高程超分辨率的工作。利用深度学习方法可以从单幅SAR图像直接推断高程信息吗？从单幅光学图像推断景深的进展和结果看<sup>[19]</sup>，原理上并不存在任何困难。深度学习从单幅SAR图像推断高程，本质上也是建立SAR图像特征与高程信息的一种映射。由于深度网络可以有效逼近任何一种函数映射关系，所以，尽管SAR图像与光学图像的成像机理不同，但从SAR图像特征到高程信息的映射函数也可以用深度网络近似。从单幅光学图像学习景深成功的另外两个操作：特征的多尺度表示和高程的局部一致性约束，SAR图像原则上也成立。所以，从单幅SAR图像在深度学习框架下直接推断高程信息，笔者觉得核心问题是“缺乏大量标注数据”。尽管“标注数据匮乏”是任何一个领域的共性问题，但SAR图像的数据匮乏现象较光学图像更为严重。

目前计算机视觉领域应对标注数据不足的基本策略是：半监督学习(semi-supervised learning)，即利用少量标注数据迭代扩大标注数据集；弱监督学习(weakly supervised learning)，即利用标注质量不高(含噪声)的数据进行学习；主动学习(active learning)，即在学习的过程中人工参与标注少量困难样本，以及模拟数据。SAR图像与波长和成像视角有关，存在相干斑噪声，比光学图像数据在数据增强方面更加困难，但笔者觉得，这也许仅仅是一个时间问题，很快含有高程信息的大量SAR标注数据集会问世。笔者觉得，在解决SAR图像标注数据匮乏问题方面，利用“仿真”和“合成”数据将是一条有效的途径。另外，鉴于目前遥感领域已有大量含有高程信息的光学影像，如何将这高程信息从光学影像迁移到SAR图像中，也是一条值得探索的途径。

总之，基于深度学习从SAR图像推断高程信息，尽管目前仍有不少难度，文献中相关报道也不多，但笔者觉得是一条值得探索且有巨大潜力和前景的技术途径。

## 6 结论

本文对SAR三维成像中的微波视觉问题进行了初步探讨。由于笔者对SAR图像处理了解不深，不妥之处在所难免，欢迎读者批评指正。关于TomoSAR中如何利用视觉语义，笔者的基本观点为：

(1) 处理框架：有必要探索基于鲁棒估计理论的“反馈式”处理框架。即先快速得到一些关于场景的粗略高程信息，然后根据这些粗略信息形成关

于场景的一些粗略三维语义约束并反馈到下一轮高程估计中。随着迭代的进行,“高程信息”和“三维场景语义”互为依托和促进,使得高程估计变得越来越准确,同时“场景语义”也变得越来越精细和可靠;

(2) 视觉语义: 场景结构基元, 如线段、面片及其位置和姿态信息, 以及物体的类别信息, 是最基本和值得优先考虑的“视觉语义”信息。这些场景结构语义信息可以通过“语义正则化”途径来描述和体现;

(3) 语义提取: 场景结构基元, 包括三维结构基元, 可以通过机器学习的途径来提取。鉴于当前机器学习缺乏“图像匹配中的外点剔除机制”, 所以如何从单幅SAR图像来提取场景结构基元以及对场景几何结构进行推断, 是一条值得探索的途径;

在结束本文之前, 笔者还想谈两点与“SAR三维成像中的视觉语义”不太关联的看法:

(1) 随着SAR成像技术的进步, SAR图像的距离和方位向分辨率会越来越高。这样单个像元内存在较多叠掩的概率也会随之减少。当叠掩次数不超过2时, 笔者觉得TomoSAR的处理技术似乎也会发生大的改变。显然, 当一个像元仅仅包含一个散射体时, 散射体对应的高程相对比较容易复原, 如理论上可以证明, 像元协方差矩阵最大特征值对应的特征向量是该散射体的投影向量。当像元包含两个散射体时, 文献[20,21]表明, 可以通过核PCA(kernel PCA)分解对应的两个最大特征值对应的特征向量来确定散射体的投影向量。这种基于PCA的方法以及其他谱分解方法, 由于计算速度快, 未来似乎应该给予必要的关注。当然, 如何估计像元的协方差矩阵本身也是一个困难的问题。

(2) TomoSAR的一个主要目标是恢复散射体的高程信息, 而恢复的高程信息往往又是一个具体应用的“中间结果”, 如对建筑物的三维重建。鉴于一般情况下, 含有多散射体的像元在整幅SAR图像中占的比例很小, 那么, 这种“耗费大量精力和时间”对单像元多散射体的努力, 对“最终目标”的实现又有多大帮助呢? 所以, 从某种程度上说, TomoSAR研究似乎也需要充分考虑具体应用(application-oriented)。当然, 从学术的观点看, 能准确恢复所有散射体的精确位置和散射性质, 永远是科学研究的一种不懈追求。

总之, SAR高分辨率和全天候的成像能力为对地观测提供了变革性的观测手段, TomoSAR的引入, 为恢复SAR高程信息提供了全新的途径。在人工智能如火如荼的今天, 微波视觉的提出, 视觉

语义的融合, 也可望为SAR快速三维成像提供有力的推动。

## 参考文献

- [1] MARR D. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information[M]. New York: W. H. Freeman, 1982.
- [2] OHZAWA I, DEANGELIS G C, and FREEMAN R D. Stereoscopic depth discrimination in the visual cortex: Neurons ideally suited as disparity detectors[J]. *Science*, 1990, 249(4972): 1037–1041. doi: [10.1126/science.2396096](https://doi.org/10.1126/science.2396096).
- [3] HAEFNER R M and CUMMING B G. Adaptation to natural binocular disparities in primate V1 explained by a generalized energy model[J]. *Neuron*, 2008, 57(1): 147–158. doi: [10.1016/j.neuron.2007.10.042](https://doi.org/10.1016/j.neuron.2007.10.042).
- [4] BRENNER E and SMEETS J B. Depth Perception[M]. WIXTED J. Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience: Sensation, Perception, and Attention. 4th ed. New York: John Wiley & Sons, 2018: 385–414.
- [5] 丁赤飏, 仇晓兰, 徐丰, 等. 合成孔径雷达三维成像——从层析、阵列到微波视觉[J]. *雷达学报*, 2019, 8(6): 693–709. doi: [10.12000/JR19090](https://doi.org/10.12000/JR19090).  
DING Chibiao, QIU Xiaolan, XU Feng, et al. Synthetic aperture radar three-dimensional imaging—from TomoSAR and array InSAR to microwave vision[J]. *Journal of Radars*, 2019, 8(6): 693–709. doi: [10.12000/JR19090](https://doi.org/10.12000/JR19090).
- [6] DEL CAMPO G M, NANNINI M, and REIGBER A. Statistical regularization for enhanced TomoSAR imaging[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 1567–1589. doi: [10.1109/JSTARS.2020.2970595](https://doi.org/10.1109/JSTARS.2020.2970595).
- [7] ZHU Xiaoxiang and BAMLER R. Tomographic SAR inversion by  $L_1$ -norm regularization—The compressive sensing approach[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2010, 48(10): 3839–3846. doi: [10.1109/TGRS.2010.2048117](https://doi.org/10.1109/TGRS.2010.2048117).
- [8] RAMBOUR C, DENIS L, TUPIN F, et al. Introducing spatial regularization in SAR tomography reconstruction[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2019, 57(11): 8600–8617. doi: [10.1109/TGRS.2019.2921756](https://doi.org/10.1109/TGRS.2019.2921756).
- [9] VIOLA P and JONES M. Rapid object detection using a boosted cascade of simple features[C]. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, USA, 2001: I–I. doi: [10.1109/CVPR.2001.990517](https://doi.org/10.1109/CVPR.2001.990517).
- [10] FISCHLER M A and BOLLES R C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography[J]. *Communications of the ACM*, 1981, 24(6): 381–395. doi: [10.1145/585846](https://doi.org/10.1145/585846).

- 1145/358669.358692.
- [11] RAMBOUR C, DENIS L, TUPIN F, *et al.* Urban surface reconstruction in SAR tomography by graph-cuts[J]. *Computer Vision and Image Understanding*, 2019, 188: 102791. doi: [10.1016/j.cviu.2019.07.011](https://doi.org/10.1016/j.cviu.2019.07.011).
- [12] LINDEBERG T. Scale-space[M]. WAH B W. Wiley Encyclopedia of Computer Science and Engineering. Hoboken: John Wiley & Sons, Inc., 2008: 2495–2504. doi: [10.1002/9780470050118.ecse609](https://doi.org/10.1002/9780470050118.ecse609).
- [13] CUI Hainan, GAO Xiang, SHNE Shuhan, *et al.* HSfM: Hybrid Structure-from-motion[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 2393–2402. doi: [10.1109/CVPR.2017.257](https://doi.org/10.1109/CVPR.2017.257).
- [14] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91–110. doi: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94).
- [15] KOLMOGOROV V and ZABIN R. What energy functions can be minimized via graph cuts?[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(2): 147–159. doi: [10.1109/TPAMI.2004.1262177](https://doi.org/10.1109/TPAMI.2004.1262177).
- [16] COSTANTE G, CIARFUGLIA T A, and BIONDI F. Towards monocular digital elevation model (DEM) estimation by convolutional neural networks - Application on synthetic aperture radar images[J]. *arXiv*: 1803.05387, 2018: 1–6.
- [17] BUDILLON A, JOHNSY A C, SCHIRINZI G, *et al.* SAR tomography based on deep learning[C]. 2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019: 3625–3628. doi: [10.1109/IGARSS.2019.8900616](https://doi.org/10.1109/IGARSS.2019.8900616).
- [18] WU Chunxiao, ZHANG Zenghui, CHEN Longyong, *et al.* Super-resolution for MIMO array SAR 3-D imaging based on compressive sensing and deep neural network[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 3109–3124. doi: [10.1109/JSTARS.2020.3000760](https://doi.org/10.1109/JSTARS.2020.3000760).
- [19] XU Dan, RICCI E, OUYANG Wanli, *et al.* Multi-scale continuous CRFs as sequential deep networks for monocular depth estimation[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 161–169. doi: [10.1109/CVPR.2017.25](https://doi.org/10.1109/CVPR.2017.25).
- [20] WANG Yuanyuan and ZHU Xiaoxiang. SAR tomography via nonlinear blind scatterer separation[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(7): 5751–5763. doi: [10.1109/TGRS.2020.3022209](https://doi.org/10.1109/TGRS.2020.3022209).
- [21] FORNARO G, VERDE S, REALE D, *et al.* CAESAR: An approach based on covariance matrix decomposition to improve multibaseline-multitemporal interferometric SAR processing[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(4): 2050–2065. doi: [10.1109/TGRS.2014.2352853](https://doi.org/10.1109/TGRS.2014.2352853).

## 作者简介



胡占义(1961–), 男, 山西繁峙人, 博士, 研究员。1985年于北方工业大学自动化系获学士学位, 1993年于比利时列日大学工学院获博士学位。1993年至今在中国科学院自动化研究所模式识别国家重点实验室工作。长期从事计算机视觉和生物视觉研究, 个人网页: <http://vision.ia.ac.cn/zh/faculty/zyhu/index.html>。

个人网页: <http://vision.ia.ac.cn/zh/faculty/zyhu/index.html>。

(责任编辑: 仇晓兰)