Interactive Endoscopy: A Next-Generation, Streamlined User Interface for Lung Surgery Navigation

Paul Thienphrapa[1], Torre Bydlon[1], Alvin Chen[1], Prasad Vagdargi[2], Nicole Varble[1], Douglas Stanton[1], and Aleksandra Popovic[1]

[1] Philips Research North America, Cambridge, MA, USA
[2] I-STAR Lab, Johns Hopkins University, Baltimore, MD, USA

Abstract. Computer generated graphics are superimposed onto live video emanating from an endoscope, offering the surgeon visual information that is hiding in the native scene—this describes the classical scenario of augmented reality in minimally invasive surgery. Research efforts have, over the past few decades, pressed considerably against the challenges of infusing a priori knowledge into endoscopic streams. As framed, these contributions *emulate perception* at the level of the surgeon expert, perpetuating debates on the technical, clinical, and societal viability of the proposition.

We herein introduce *interactive endoscopy*, transforming passive visualization into an interface that allows the surgeon to label noteworthy anatomical features found in the endoscopic video, and have the virtual annotations remember their tissue locations during surgical manipulation. The streamlined interface combines vision-based tool tracking and speech recognition to enable interactive selection and labeling, followed by tissue tracking and optical flow for label persistence. These discrete capabilities have matured rapidly in recent years, promising technical viability of the system; it can help clinicians offload the cognitive demands of visually deciphering soft tissues; and supports societal viability by engaging, rather than emulating, surgeon expertise. Through a video-assisted thoracotomy use case, we develop a proof-of-concept to improve workflow by tracking surgical tools and visualizing tissue, while serving as a bridge to the classical promise of augmented reality in surgery.

Keywords: Interactive Endoscopy · Lung surgery · VATS · Augmented Reality · Human-Computer Interaction

1 Introduction and Motivation

Lung cancer is the deadliest form of cancer worldwide, with 1.6 million new diagnoses and 1.4 million deaths each year, more than cancers of the breast, prostate, and colon—the three next most prevalent cancers—combined. In response to this epidemic, major screening trials have been enacted including the Dutch/Belgian NELSON trial, the US NLST, and Danish trials. These studies

found that proactive screening using low dose computed tomography (CT) can detect lung cancer at an earlier, treatable stage at a rate of 71%, leading to a 20% reduction in mortality [2]. This prompted Medicare to reimburse lung cancer screening in 2015 and with that, the number of patients presenting with smaller, treatable tumors was expected to rise dramatically. The projected increase was observed within the Veterans Health Administration [17], and while this population bears a heightened incidence of lung cancer due to occupational hazards, the need to optimize patient care was foretold.

Surgical resection is the preferred curative therapy due to the ability to remove units of anatomy that sustain the tumor, as well as lymph nodes for staging. Most of the 100,000 surgical resections performed in the US annually are minimally invasive, with 57% as video-assisted thoracoscopic surgery (VATS) and 7% as robotic surgery [1]. Anatomically, the lung follows a tree structure with airways that root at the trachea and narrow as they branch towards the ribcage; blood vessels hug the airways and join them at the alveoli, or air sacs, where oxygen and carbon dioxide interchange. Removing a tumor naturally detaches downstream airways, vessels, and connective lung tissue, so tumor location and size prescribe the type of resection performed. Large or central tumors are removed via pneumonectomy (full lung) or lobectomy (full lobe), while small or peripheral tumors may be "wedged" out. Segmentectomy, or removal of a sub-lobar segment, is gaining currency because the procedure balances disease removal with tissue preservation; and because the trend towards smaller, peripheral tumors supports it.

2 Background

In an archetypal VATS procedure, the surgeon examines a preoperative CT; here the lung is inflated. They note the location of the tumor, relative to adjacent structures. Now under the thoracoscope, the lung is collapsed, the tumor invisible. The surgeon roughly estimates the tumor location. They look for known structures; move the scope; manipulate the tissue; reveal a structure; remember it. They carefully dissect around critical structures [14]; discover another; remember it. A few iterations and their familiarity grows. They mentally align new visuals with their knowledge, experience, and the CT. Thusly, they converge on an inference of the true tumor location.

The foregoing exercise is cognitively strenuous, time consuming, yet merely a precursor to the primary task of tumor resection. It is emblematic of endoscopic surgery in general and of segmentectomy in particular, as the surgeon continues to mind critical structures under a limited visual field [18]. Consequently, endoscopic scenes are difficult to contextualize in isolation, thereby turning the lung into a jigsaw puzzle in which the pieces may deform, and must be memorized. Indeed, surgeons routinely retract the scope or zoom out to construct associations and context. Moreover, the lung appearance may not be visually distinctive nor instantly informative, further intensifying the challenges, and thus the inefficiencies, of minimally invasive lung surgery.

The research community has responded vigorously, imbuing greater context into endoscopy using augmented reality [5, 24] by registering coherent anatomical models onto disjoint endoscopic snapshots. For example, Puerto-Souza et al. [25] maintain registration of preoperative images to endoscopy by managing tissue anchors amidst surgical activity. Du et al. [11] combine features with lighting, and Collins et al. [9], texture with boundaries, to track surfaces in 3D. Lin et al. [19] achieve surface reconstruction using hyperspectral imaging and structured light. Simultaneous localization and mapping (SLAM) approaches have been extended to handle tissue motion [23] and imaging conditions [21] found in endoscopy. Stereo endoscopes are used to reconstruct tissue surfaces with high fidelity, and have the potential to render overlays in 3D or guide surgical robots [29, 32]. Ref. [22] reviews the optical techniques that have been developed for tissue surface reconstruction.

In recent clinical experiments, Chauvet et al. [8] project tumor margins onto the surfaces of *ex vivo* porcine kidneys for partial nephrectomy. Liu et al. [20] develop augmented reality for robotic VATS and robotic transoral surgery, performing preclinical evaluations on ovine and porcine models, respectively, in elevating the state of the art. These studies uncovered insights on registering models to endoscopy and portraying these models faithfully. However, whether pursuing a clinical application or technical specialty, researchers have faced a timeless obstacle: tissue deformation. Modeling deformation is an ill-posed problem, and this coinciding domain has likewise undergone extensive investigation [28]. In the next section, we introduce an alternative technology for endoscopy that circumvents the challenges of deformable modeling.

3 Interactive Endoscopy

3.1 Contributions

The surgeon begins a VATS procedure as usual, examining the CT, placing surgical ports, and estimating the tumor location under thoracoscopy. They adjust both scope and lung in search of known landmarks, the pulmonary artery for instance. Upon discovery, now under the proposed interactive endoscopy system (Fig. 1, from a full *ex vivo* demo), they point their forceps at the target and verbally instruct, "Mark the pulmonary artery." An audible chime acknowledges, and a miniature yet distinctive icon appears in the live video at the forceps tip, accompanied by the semantic label. The surgeon continues to label the anatomy and as they move the scope or tissue, the virtual annotations follow.

The combination of a limited visual field and amorphous free-form tissue induces the surgeon to perform motions that are, in technical parlance, computationally intractable—it is the practice of surgery itself that postpones surgical augmented reality ever farther into the future. In that future, critical debates on validation and psychophysical effects await, yet the ongoing challenges of endoscopic surgery have persisted for decades. The present contribution transforms endoscopy from a passive visualization tool to an interactive interface for labeling live video. We recast static preoperative interactivity from Kim et al. [16] into

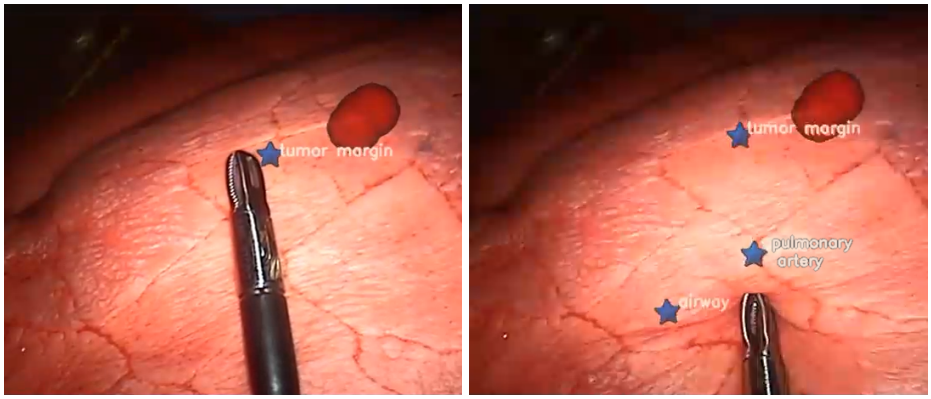

Fig. 1. Interactive endoscopy annotation system on an *ex vivo* porcine lung; the red simulated tumor is part of the live demonstration. (*Left*) The surgeon points a tool at a feature of interest then speaks the desired label, "tumor margin". (*Right*) Multiple labels are tracked as the surgeon manipulates the tissue. Note that the system is non-disruptive to existing workflows and requires no setup.

an intraoperative scheme; and repurpose OCT review interactivity from Balicki et al. [4] to provide live spatial storage for the expert's knowledge. We show how the system can help surgeons through a cognitively strenuous and irreproducible exploration routine, examine circumstances that would enable clinical viability, and discuss how the approach both complements and enables augmented reality in surgery, as envisioned a generation ago (Fuchs et al., 1998 [15]).

3.2 Key Components

For the proposed interactive endoscopy system, the experimental setup and usage scenario are pictured in Fig. 2. Its key components include (1) vision-based tool tracking, (2) a speech interface, and (3) persistent tissue tracking. While these discrete capabilities have been historically available, they have undergone marked improvement in recent years due to the emergence of graphical processing units (GPUs), online storage infrastructure, and machine learning. A system capitalizing on these developments has the potential to reach clinical reliability in the near future. While these technologies continue to evolve rapidly, we construct a proof-of-concept integration of agnostic building blocks as a means of assessing baseline performance.

Tool Tracking. Upon discovering each landmark, the surgeon points out its location to the system. A workflow-compatible pointer can be repurposed from a tool already in use, such as a forceps, by tracking it in the endoscope. We use a hierarchical heuristic method (Fig. 3) with similar assumptions as in [10] that thoracic tools are rigid, straight, and of a certain hue. The low-risk demands of the pointing task motivates our simple approach: 2D tool tracking can reliably

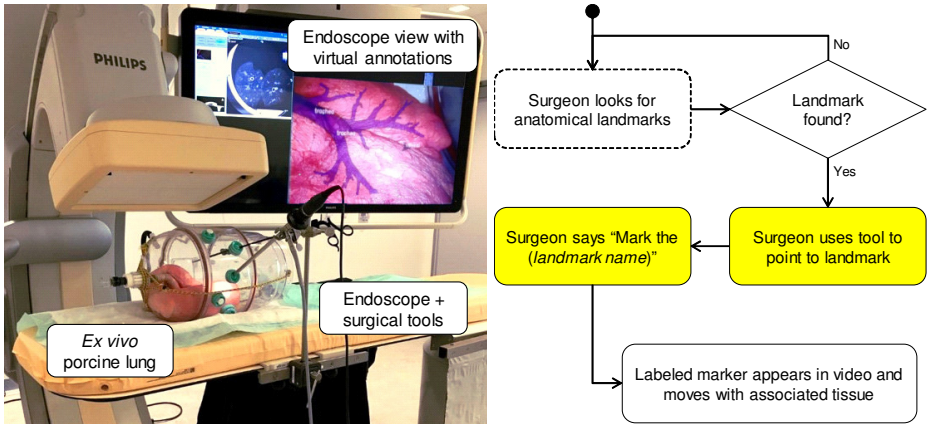

Fig. 2. (*Left*) Experimental setup. Minimal instrumentation beyond that of standard VATS is required, primarily a microphone and a computer (the C-arm is presently unused). (*Right*) Workflow for applying a label. Pointing and verbal annotation (yellow) are likewise minimal steps.

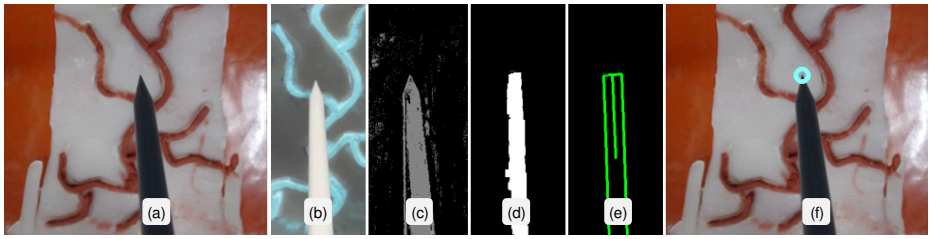

Fig. 3. Tool tracking pipeline: (a) Original (b) HSV threshold (c) Foreground learning using MoG (d) Foreground mask (e) Contour detection (f) Tip detection.

map 3D surface anatomy due to the projective nature of endoscopic views. Our *ex vivo* tests indicate 2D tip localization to within 1.0 mm 92% of the time that the tool points to a new location, and more advanced methods [6, 27] suggest that clinical-grade tool tracking is well within reach.

Speech Interface. Pointing the forceps at a landmark, the surgeon uses speech to generate a corresponding label. This natural, hands-free interface is conducive to workflow and sterility, as previously acknowledged in the voice-controlled AESOP endoscope robot [3]. Recognition latency and accuracy were at the time prohibitive, but modern advances have driven widespread use. The proliferation of voice-controlled virtual assistants (e.g., Alexa) obliges us to revisit speech as a surgical interface.

We use Google Cloud Speech-to-Text in our experiments. The online service allows the surgeon to apply arbitrary semantic labels; offline tools or preset

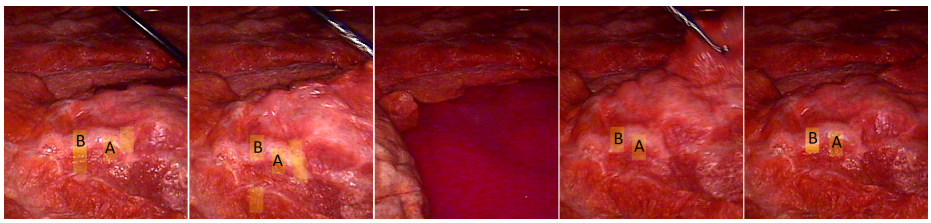

Fig. 4. Persistent tissue tracking through folding and unfolding. SURF features (A and B) are labeled so that annotations can be maintained though surgical manipulation.

vocabularies may be preferred in resource-constrained settings. Qualitatively, we observed satisfactory recognition during demos using a commodity wireless speaker-microphone, and this performance was retained in a noisy exhibition room by switching to a wireless headset, suggesting that the clinical benefits of speech interfaces may soon be realizable.

Persistent Tissue Tracking. After the surgeon creates a label, the system maintains its adherence by encoding the underlying tissue and tracking it as the lung moves, deforms, or reappears in view following an exit or occlusion. This provides an intuition of the lung state, with similar issues faced in liver surgery. The labeling task asks that arbitrary patches of tissue be persistently identified whenever they appear—a combination of online detection and tracking which for endoscopy is well served by SURF and optical flow [12]. SURF can identify tissue through motion and stretching 83% and 99% of the time respectively [30]. In *ex vivo* experiments, uniquely identified features could be recovered successfully upon returning into view so long as imaging conditions remain reasonably stable, as illustrated in Fig. 4.

Labels should be displayed, at minimum, when the tissue is at rest, and modern techniques in matching sub-image elements [13, 33] show promise in overcoming the challenges of biomedical images [26]. Approaches such as FlowNet can then be used to track moving tissue and enhance the realism of virtual label adherence. In short, there is a new set of tools to address traditional computer vision problems in endoscopy.

3.3 Capturing Surgeon Expertise

Interactive endoscopy ties maturing technologies together into a novel application with forgiving performance requirements, paving the way to the level of robustness needed for clinical use. The simplicity of the concept belies its potential to alleviate cognitive load, which can impact both judgment and motor skills [7]. When the surgeon exerts mental energy in parsing what they see, the system lets them translate that expertise directly onto the virtual surgical field. This mitigates redundant exploration, and the visibility of labels can help them infer context more readily.

In fact, many surgeons already use radiopaque, tethered (iVATS), dye [31], and ad hoc electrocautery markers to aid localization prior to or during surgery. These varied practices introduce risk and overhead, whereas virtual markers are easy to use and provide a reason for use, potentially bridging gaps between clinical practices and supporting technology. Moreover, surgeon engagement with technology has a broader implication: digitization of the innards of surgery, which has been a black box. Digital labels offer a chance to capture semantic, positional, temporal, visual, and procedural elements of surgery, forming a statistical basis for understanding—and anticipating—surgical acts at multiple scales. This, in turn, can help make augmented reality a clinical reality.

4 Conclusions

The promise of augmented reality in surgery has been tempered by challenges such as soft tissue deformation, and efforts to overcome this timeless adversary has inadvertently suspended critical debates on the role of augmented perception in medicine altogether. We present, as a technological bridge, a streamlined user interface that allows surgeons to tag the disjoint views that comprise endoscopic surgery. These virtual labels persist as the organ moves, so surgeons can potentially manage unfamiliar tissue more deterministically. This approach embraces the finiteness of human cognition and alleviates reliance on cognitive state, capturing expert perception and judgment without attempting to emulate it. We design a minimal feature set and a choice architecture with symmetric freedom to use or not, respecting differences between surgeons. Our baseline system demonstrates promising performance in a lab setting, while rapid ongoing developments in the constituent technologies offer a path towards clinical robustness. These circumstances present the opportunity for surgeons to change surgery, without being compelled to change.

References

1. Healthcare Cost and Utilization Project, https://hcupnet.ahrq.gov/#setup
2. Reduced lung-cancer mortality with low-dose computed tomographic screening. New England Journal of Medicine **365**(5), 395–409 (2011)
3. Allaf, M.E., Jackman, S.V., Schulam, P.G., Cadeddu, J.A., Lee, B.R., Moore, R.G., Kavoussi, L.R.: Laparoscopic visual field. Surg. Endosc. **12**(12), 1415–1418 (1998)
4. Balicki, M., Richa, R., Vagvolgyi, B., Kazanzides, P., Gehlbach, P., Handa, J., Kang, J., Taylor, R.: Interactive OCT annotation and visualization for vitreoretinal surgery. In: Augmented Environments for Computer-Assisted Interventions (2013)
5. Bernhardt, S., Nicolau, S.A., Soler, L., Doignon, C.: The status of augmented reality in laparoscopic surgery as of 2016. Med. Image Anal. **37**, 66–90 (2017)
6. Bodenstedt, S., et al.: Comparative evaluation of instrument segmentation and tracking methods in minimally invasive surgery (2018)
7. Carswell, C.M., Clarke, D., Seales, W.B.: Assessing mental workload during laparoscopic surgery. Surg. Innov. **12**(1), 80–90 (2005)

8. Chauvet, P., Collins, T., Debize, C., Novais-Gameiro, L., Pereira, B., Bartoli, A., Canis, M., Bourdel, N.: Augmented reality in a tumor resection model. Surg. Endosc. **32**(3), 1192–1201 (2018)
9. Collins, T., Bartoli, A., Bourdel, N., Canis, M.: Robust, real-time, dense and deformable 3D organ tracking in laparoscopic videos. In: Medical Image Computing and Computer-Assisted Intervention. pp. 404–412 (2016)
10. Doignon, C., Nageotte, F., de Mathelin, M.: Segmentation and guidance of multiple rigid objects for intra-operative endoscopic vision. In: Vidal, R., et al. (eds.) Dynamical Vision. pp. 314–327. Springer Berlin Heidelberg (2007)
11. Du, X., Clancy, N., Arya, S., Hanna, G.B., Kelly, J., Elson, D.S., Stoyanov, D.: Robust surface tracking combining features, intensity and illumination compensation. Int. J. Comput. Assist. Radiol. Surg. **10**(12), 1915–1926 (2015)
12. Elhawary, H., Popovic, A.: Robust feature tracking on the beating heart for a robotic-guided endoscope. Int. J. Med. Robot. Comput. Assist. Surg. **7**(4) (2011)
13. Fischer, P., Dosovitskiy, A., Brox, T.: Descriptor matching with convolutional neural networks: A comparison to SIFT (2014)
14. Flores, R.M., et al.: Video-assisted thoracoscopic surgery (VATS) lobectomy: Catastrophic intraoperative complications. J. Thorac. Cardiovasc. Surg. **142**(6), 1412–1417 (2011)
15. Fuchs, H., et al.: Augmented reality visualization for laparoscopic surgery. In: Medical Image Computing and Computer-Assisted Intervention. pp. 934–943. Springer Berlin Heidelberg (1998)
16. Kim, J.H., Bartoli, A., Collins, T., Hartley, R.: Tracking by detection for interactive image augmentation in laparoscopy. In: Dawant, B.M., et al. (eds.) Biomedical Image Registration. pp. 246–255. Springer Berlin Heidelberg (2012)
17. Kinsinger, L.S., et al.: Implementation of lung cancer screening in the Veterans Health Administration. JAMA Internal Medicine **177**(3), 399–406 (2017)
18. Lee, C.Y., Chan, H., Ujiie, H., Fujino, K., Kinoshita, T., Irish, J.C., Yasufuku, K.: Novel thoracoscopic navigation system with augmented real-time image guidance for chest wall tumors. Ann. Thorac. Surg. **106**(5), 1468–1475 (2018)
19. Lin, J., Clancy, N.T., Qi, J., Hu, Y., Tatla, T., Stoyanov, D., Maier-Hein, L., Elson, D.S.: Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks. Med. Image Anal. **48**, 162–176 (2018)
20. Liu, W.P., Richmon, J.D., Sorger, J.M., Azizian, M., Taylor, R.H.: Augmented reality and CBCT guidance for transoral robotic surgery. J. Robot. Surg. **9**(3), 223–233 (2015)
21. Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., Montiel, J.M.M.: Live tracking and dense reconstruction for handheld monocular endoscopy. IEEE Trans. Med. Imaging **38**(1), 79–89 (2019)
22. Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., Kolb, A., Rodrigues, M., Sorger, J., Speidel, S., Stoyanov, D.: Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. Med. Image Anal. **17**(8), 974–996 (2013)
23. Mountney, P., Yang, G.Z.: Motion compensated SLAM for image guided surgery. In: Medical Image Computing and Computer-Assisted Intervention. pp. 496–504. Springer Berlin Heidelberg (2010)
24. Nicolau, S., Soler, L., Mutter, D., Marescaux, J.: Augmented reality in laparoscopic surgical oncology. Surg. Oncol. **20**(3), 189–201 (2011)

25. Puerto-Souza, G.A., Cadeddu, J.A., Mariottini, G.L.: Toward long-term and accurate augmented-reality for monocular endoscopic videos. IEEE Trans. Biomed. Eng. **61**(10), 2609–2620 (2014)
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention. pp. 234–241. Springer International Publishing (2015)
27. Shvets, A.A., Rakhlin, A., Kalinin, A.A., Iglovikov, V.I.: Automatic instrument segmentation in robot-assisted surgery using deep learning. In: IEEE Int. Conf. on Machine Learning and Applications (ICMLA). pp. 624–628 (2018)
28. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE Trans. Med. Imaging **32**(7), 1153–1190 (2013)
29. Stoyanov, D., Scarzanella, M.V., Pratt, P., Yang, G.Z.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: Medical Image Computing and Computer-Assisted Intervention. pp. 275–282 (2010)
30. Thienphrapa, P., Bydlon, T., Chen, A., Popovic, A.: Evaluation of surface feature persistence during lung surgery. In: BMES Annual Meeting. Atlanta, GA (2018)
31. Willekes, L., Boutros, C., Goldfarb, M.A.: VATS intraoperative tattooing to facilitate solitary pulmonary nodule resection. J. Cardiothorac. Surg. **3**(1), 13 (2008)
32. Yip, M.C., Lowe, D.G., Salcudean, S.E., Rohling, R.N., Nguan, C.Y.: Tissue tracking and registration for image-guided surgery. IEEE Trans. Med. Imaging **31**(11), 2169–2182 (2012)
33. Zagoruyko, S., Komodakis, N.: Learning to compare image patches via CNNs. In: IEEE Conf. on Computer Vision and Pattern Recognition. pp. 4353–4361 (2015)