Automatic 2D-3D Vision Based Assessment of the Attitude of a Train Pantograph

Erika Di Stefano, Emanuele Ruffaldi, Carlo Alberto Avizzano

***Abstract*— In this paper we propose an automatic visual based technique, integrated in a wayside monitoring system for train inspection, that allows to assess the attitude of the metal bow of a pantograph by combining a colour image captured by an RGB digital camera and a pointcloud built from a range sensor scan. An efficient and fast template-matching procedure allows to detect the pantograph in the scene and associate a matching attitude, searching for the most similar model present in a database. The record of templates belonging to the database exploits a virtual rendering environment that allows to optimize the training stage in terms of computational load and time. During actual inspection the RGB image and pointcloud of the pantograph are opportunely processed and aligned to the same reference frame. After the preliminary template-matching step, the pointcloud is augmented with the virtual model of the matched template and the attitude angular values are refined by applying the iterative closest point (ICP) algorithm between the real object and the virtual one, with the aim of reducing eventual residual errors.**

I. INTRODUCTION

Condition based maintenance is nowadays an increasing common practice in international railways. Manual inspection operations are usually scheduled on regular basis plans that depend on time and travel distance. This type of verification usually takes much time and sometimes requires the interruption of train operation which also implies significant costs for the railway companies. This is the reason for which much research focuses on the realization of automatic monitoring systems that support and optimize the diagnosis process in terms of costs and time. The pantograph is one of the most monitored components among train parts, due to its fundamental role in the train running and, at the same time, to its tendency of wear and high probability of suffering damage. Despite its robust construction structure, the pantograph is indeed subject to high mechanical stress and accidental strokes with surrounding obstacles met during the train travel, such as tree branches, are not uncommon. As a consequence, it can undergo deformations, misalignments and attitude changes that could alter the electrical transmission and may result in a system malfunction. A prompt detection of the alterations is therefore essential to proceed to the maintenance and avoid further damage.

Early attempts to predict failure using automated techniques in the transport sector were proposed at the end of eighties. In [1] Betts focuses mostly on indirect measurements of pantograph-catenary contacts. In [2], Landi proposes the use of the Hough transform as a means to monitor the linearity of the strip contact surface recorded with an infra-red camera. In [3], Aydin proposes the use of a digital image taken from a side camera combined with the use of gaussian mixture models (GMM) to identify anomalies at the contacts. In [4] Jarzebowicz integrates a 3D vision technique, based on laser triangulation to measure wear and damages on the carbon strips.

The most recent commercial systems are the Chinese SJ Wayside inspection system, the U.S. Duostech APIS system, the DK PantoInspect system and the PantoBot-Pavisys system [5]. These systems are mostly focused on the detection of surface material defects, such as wear of the carbon strips, presence of chipping or foreign material and material damage.

To our knowledge, none of these systems provides a complete assessment of major fault conditions, such as: major geometrical changes (such as consistent material loss), 6 degrees of freedom misalignments issues including frontal, lateral and upwards as well as yaw, pitch and roll distortions, historical analysis of pose changes.

The present work is being carried out between Scuola Superiore Sant'Anna and TRENITALIA S.p.A. cooperation. The cooperation aims at developing a full set of wayside inspection tools for monitoring the health status of trains. At present the concept of the portal and positioning of the sensors is still under development and preliminary experiments are being carried out in laboratory facilities and wayside the maintenance plants, where it would be easier to manage temporary set of cameras and robots.

The vision system presented in this paper is one of the algorithms currently under development for the assessment of the pantograph attitude. We developed an automatic visual based inspection system that allows to compute the complete attitude from the combination of an RGB image and a pointcloud. The attitude assessment is based on a template-matching algorithm, followed by the ICP technique [6]. The introduction of virtual models allows to optimize the computational load and time of the training stage on one hand and to reduce the error of the final result on the other. The paper is structured as follows: In section II we illustrate the pointcloud and RGB image alignment process, in section III we explain the model based template matching technique, in section IV we explain the matching and iterative closest point refinement technique, section V shows the experimental results.

II. RGB IMAGE AND POINTCLOUD ALIGNMENT

Since the RGB camera and the range sensor may be positioned in different ways, an accurate calibration procedure

*PERCeptual RObotics Laboratory, Scuola Superiore Sant'Anna

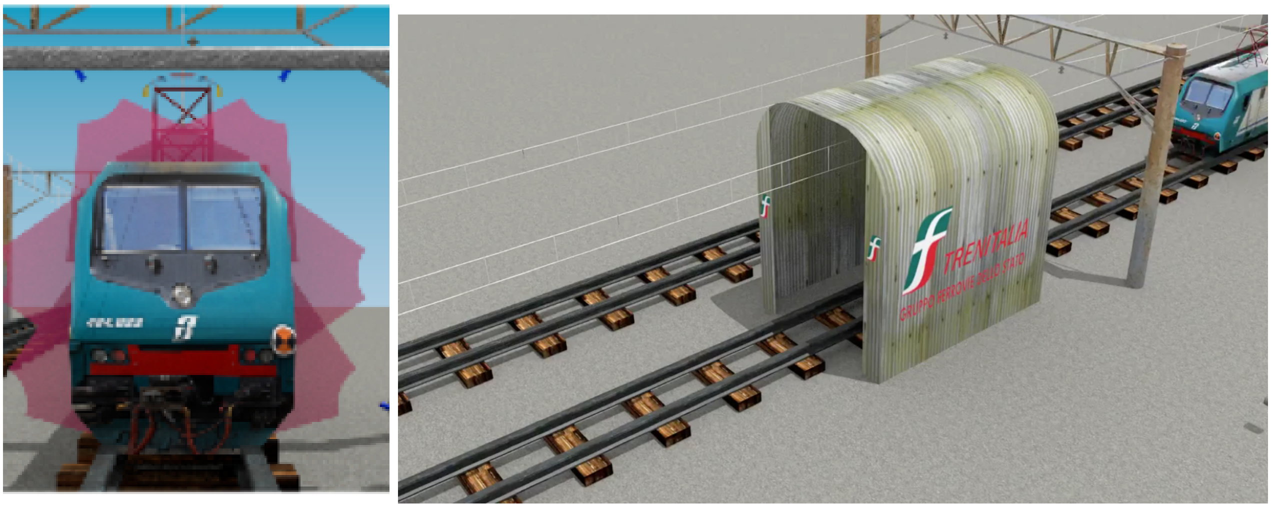

Fig. 1. The concept of the wayside monitoring portal. Two laser camera from the top will capture pantograph and train roof data. Structured light camera on the side will capture the carriage information.

is required to compute the relative pose between them. The knowledge of the relative pose is essential to align the RGB image and the pointcloud data and extract the local features coherently.

Let's denote with $\{D\}$ the laser frame and with $\{C\}$ the RGB camera frame and indicate respectively with ${}^{D}\boldsymbol{p}$ and ${}^{C}\boldsymbol{p}$ the 3D coordinates of a world point with respect to the laser and the camera frame. Given the relative rotation ${}^{C}\boldsymbol{R}_{D}$ and translation ${}^{C}\boldsymbol{T}_{D}$ between the two frames, the world point expressed in $\{D\}$ can be mapped to $\{C\}$ as follows :

$$ {}^{C}\boldsymbol{p} = {}^{C}\boldsymbol{R}_{D}\,{}^{D}\boldsymbol{p} + {}^{C}\boldsymbol{T}_{D} \tag{1} $$

and projected, according to perspective projection and the pinhole camera model [7], onto the camera image plane. Denoted with ${}^{C}x$, ${}^{C}y$ and ${}^{C}z$ the 3D coordinates of the point with respect to the camera frame, the corresponding pixel coordinates u and v are obtained from equation (2):

$$ \begin{cases} u = \frac{{}^{C}x}{{}^{C}z} f_x + c_x \\ v = \frac{{}^{C}y}{{}^{C}z} f_y + c_y \end{cases} \tag{2} $$

where f_x, f_y, c_x and c_y are the camera intrinsic parameters. By taking the pixels corresponding to each remapped point after applying equations (1 - 2), we obtain the RGB image aligned to the pointcloud and can extract the features needed for the matching.

III. TEMPLATE BASED MODEL MATCHING

The literature is full of works that deal with pose estimation of rigid bodies. Many of them are based on geometrical features, such as points or lines [8], [9], [10]. In these cases the points/features extracted must be identified uniquely and, however overall information about the body may be lost (single features can be extracted despite loss or damages of some of its parts). On the contrary, a template-based method carries out a comparison of the scene with a reference model and fails in case of deviation of the real object from the

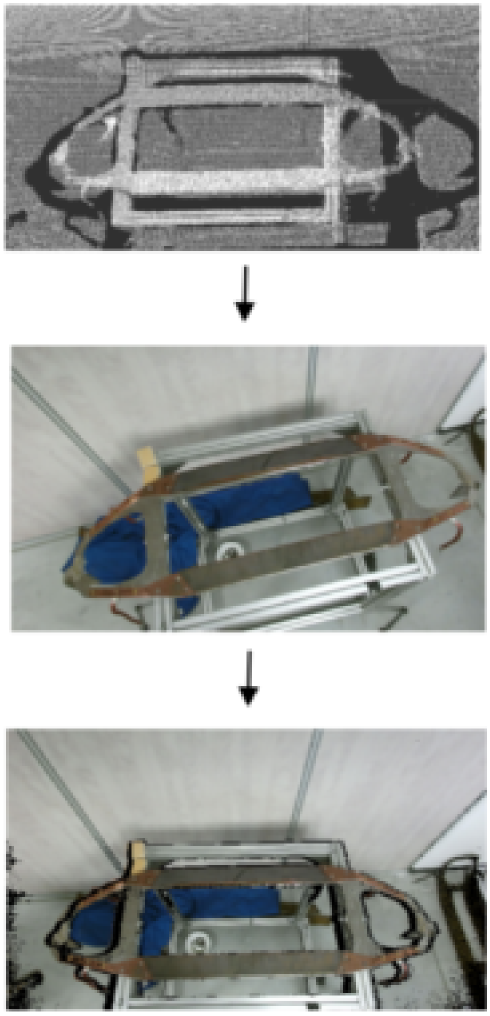

Fig. 2. Pointcloud, RGB image and resulting aligned RGB image.

intact model. The one employed in this work is described in [11], since we considered it most suitable for the data of our interest and it allows us to exploit the coexistence of 2D and 3D data in an optimized way. In short, the algorithm is based on the extraction of local features (quantized colour gradients and surface normals orientations) from RGB and depth data and their comparison with previously stored templates. Comparison is carried out through simple binary strings obtained from the processing of the features and thus the method is very fast and efficient. A matching occurs, when an adjustable similarity score threshold is reached. On the other hand, the storage of templates, as previously mentioned, requires the acquisition of numerous colour images and pointclouds of the pantograph from which the visual features have to be extracted and templates have to be created, so as to match the attitude seen in the scene with one of the recorded ones. Manual acquisition requires time and highly accurate attitude recordings. Hence, to accelerate the procedure, we introduced a virtual reality module based on the OpenGL library [12] that automatically generates a wide set of views of the object from its CAD model. During the training process, we assign different attitudes to the CAD model and a virtual rendering process projects the CAD model onto the image plane of a virtual camera. In this way we create a wide set of synthetic RGB images and pointclouds depicting the pantograph in different attitudes. Figures 5 and 4 illustrate the basic concept of the rendering process. The image projections are made at a work-distance, and for different values of the roll-pitch-yaw angles (with a 2^o step). At the same time, the feature training system takes the synthetic RGB and depth data to compute the visual features and create a template that labels a specific attitude configuration.

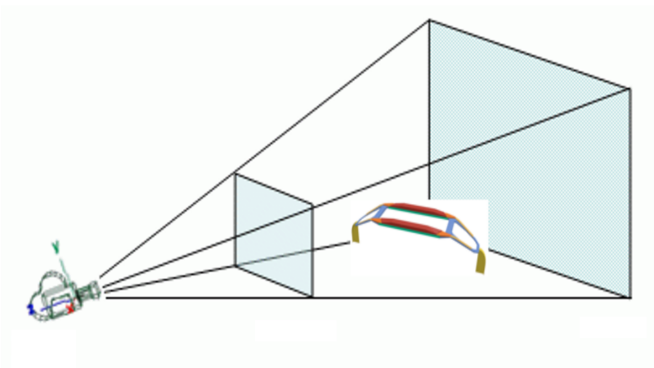

Fig. 3. Basic concept of virtual rendering process. The CAD model is projected onto the image of a virtual camera.

IV. MATCHING AND ICP REFINEMENT

In a first step we roughly search for the most similar templates that match the features computed in the RGB image and pointcloud data. Since the matching is based on a similarity threshold, more templates that match the current scene can be selected. As a similarity score is associated to each matched template, we compute the matched pose as

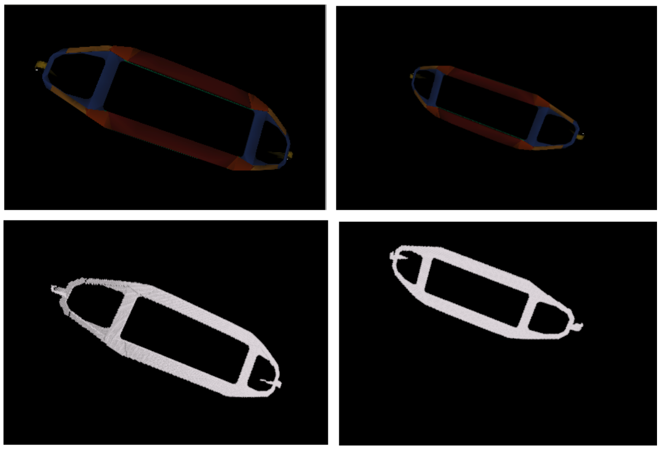

Fig. 4. RGB image and pointcloud resulting from the rendering of the CAD model of the metal bow of an FS 52/92 pantograph, projected in different poses with respect to the virtual camera frame.

the weighted average of the selected ones, where the weights are the respective similarity scores and assuming an angular range that rules out the singularity cases.

$$\begin{cases} \psi = \frac{\sum_i s_i \psi_i}{s_i} \\ \phi = \frac{\sum_i s_i \phi_i}{s_i} \\ \theta = \frac{\sum_i s_i \theta_i}{s_i} \end{cases} \tag{3}$$

where ψ, ϕ and θ are the body's yaw, pitch and roll angles and s_i and and ψ_i, ϕ_i and θ_i respectively the similarity score and attitude angles associated to the i-th matched template. Next, the virtual pointcloud of the matched template is projected onto the scene, the real pantograph pointcloud is extracted as a cluster from the entire cloud and the two clouds are more finely aligned through the iterative closest point technique. The rotation matrix $\boldsymbol{R}(\psi_f, \phi_f, \theta_f)$ encoding the body's attitude is eventually corrected by the further rotation matrix $\boldsymbol{R}(\psi_r, \phi_r, \theta_r)$ that aligns the template virtual cloud to the real one:

$$\boldsymbol{R}(\psi_f, \phi_f, \theta_f) = \boldsymbol{R}(\psi, \phi, \theta)\, \boldsymbol{R}(\psi_r, \phi_r, \theta_r) \tag{4}$$

V. EXPERIMENTAL RESULTS

Since the laser scanner and high resolution camera expected to be used for the purpose weren't available yet, we carried out a set of preliminary experiments by employing an uncalibrated KINECT for XBOX ONE. The sensor furnishes both a pointcloud, as a matrix of 3D points expressed in the camera frame, and an aligned RGB image, which in case of uncertain intrinsic parameters and relative pose between the RGB and IR camera can be subject to alignment errors. Moreover, as shown in [13] the sensor's accuracy decreases with the distance and the pointcloud is affected by measurement noise

In order to validate our visual based attitude assessment system, we positioned the pantograph bow on a metal support and manually and randomly rotated it in a way to record different attitudes. Our results were compared with those

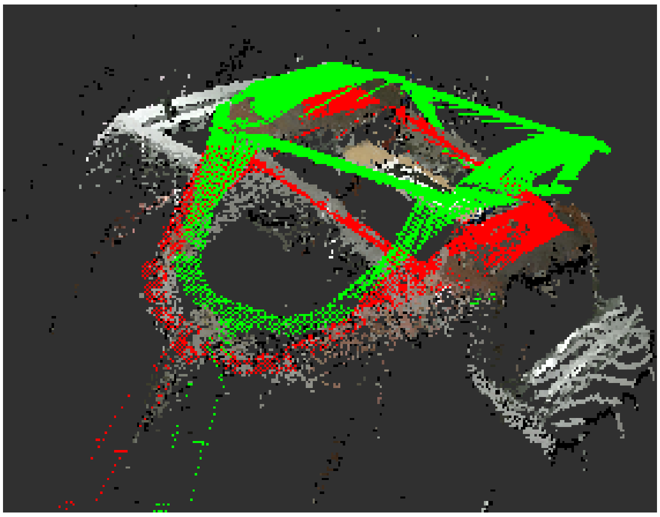

Fig. 5. Example case of when ICP corrects the template-matching residual error. The green cloud is the virtual cloud of the matched template, while the red cloud is the one resulting after ICP application.

obtained by an OptiTrack motion capture system (see Figure 6).

Fig. 6. System of Data Acquisition: Optitrack and Kinect for Xbox One.

From the results, we notice that the angular errors are significantly reduced after ICP alignment. Tables I and II show the minimum, maximum and average error values obtained first considering only the matched template and then refining with ICP.

In Figure (7) the red pointcloud is the virtual pointcloud obtained from the CAD model and projected onto the scene according to the position and attitude resulting computed by our algorithm. The virtual pointcloud overlaps the real pantograph bow in the pointcloud. The green points on the aligned RGB image are the matching points found by the template-matching algorithm. A further window shows the CAD model rendered according to the computed pose. Figure (8) shows the real and overlapped virtual pointclouds from different points of view.

TABLE I

MINIMUM, MAXIMUM AND AVERAGE ERROR VALUES (IN DEGREES) DERIVING FROM THE TEMPLATE-MATCHING.

	YAW	PITCH	ROLL
MIN	0.9521	3.5437	5.9381
MAX	5.8532	9.9304	12.6316
AVERAGE	2.7836	5.7219	8.8982

TABLE II

MINIMUM, MAXIMUM AND AVERAGE ERROR VALUES (IN DEGREES) AFTER ICP APPLICATION.

	YAW	PITCH	ROLL
MIN	0.0484	0.8508	0.3363
MAX	5.4081	5.4430	5.5318
AVERAGE	2.3245	2.5902	2.5581

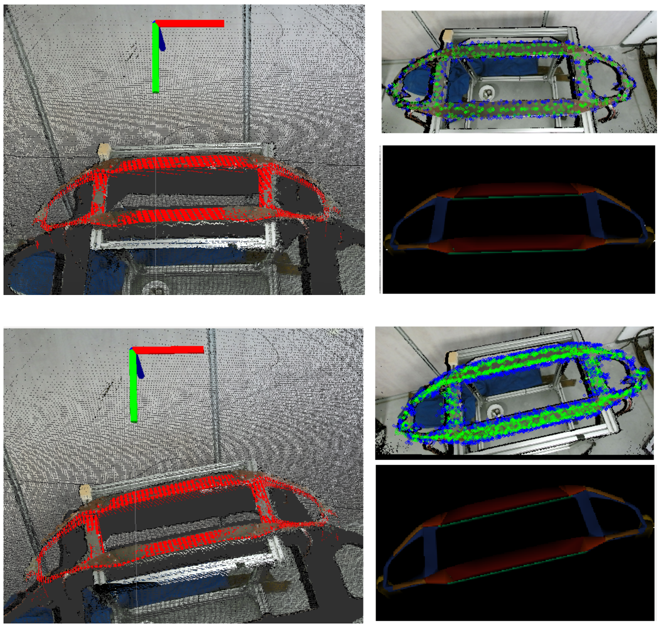

Fig. 7. Result of template matching and ICP in two of the arbitrary attitude placements. The left image shows the virtual model pointcloud (in red) overlapped on the real pointcloud after the estimation process. The right image on the top shows the matching points (in green) extracted from the RGB image, while the one at the bottom shows the virtual model RGB image rendered according to the estimated attitude. The frame on the 3D image represents the camera reference frame.

VI. CONCLUSIONS AND FUTURE WORK

We described an automatic visual-based technique for the assessment of the attitude of a train pantograph bow. Our system is able to compute the attitude from the pointcloud and the RGB image of the pantograph bow by means of a template-matching algorithm, followed by an iterative closest point refinement. The first one is based on local visual features, which are extracted from the current data and compared to a set of previously stored templates, built

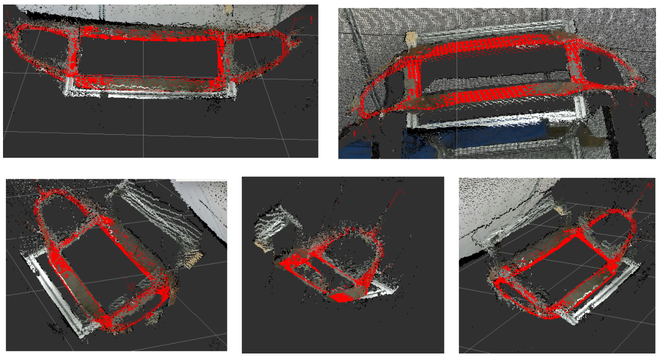

Fig. 8. Overlapped model virtual pointcloud seen from different viewpoints.

from virtual models, in order to find best matches. The subsequent ICP alignment process serves to improve the found attitude and is applied between the real pointcloud of the pantograph bow and the virtual pointcloud of the matched template . Experimental results show that the system has good performance and that the combination of template-matching and ICP leads to low maximum errors.

Future work will concern the setup of the actual sensors and experimental tests carried out by evaluating also the robustness of our algorithm to mis-alignments.

References

[1] A. Betts, J. Hall, and P. Keen, "Condition monitoring of pantographs," in *Main Line Railway Electrification, 1989., International Conference on*. IET, 1989, pp. 129–133.

[2] A. Landi, L. Menconi, and L. Sani, "Hough transform and thermo-vision for monitoring pantograph-catenary system," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 220, no. 4, pp. 435–447, 2006.

[3] I. Aydin, M. Karakose, and E. Akin, "A robust anomaly detection in pantograph-catenary system based on mean-shift tracking and foreground detection," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on*. IEEE, 2013, pp. 4444–4449.

[4] L. Jarzebowicz and S. Judek, "3d machine vision system for inspection of contact strips in railway vehicle current collectors," in *Applied Electronics (AE), 2014 International Conference on*. IEEE, 2014, pp. 139–144.

[5] M. Sacchi, S. Cagnoni, D. Spagnoletti, L. Ascari, G. Zunino, and A. Piazzi, "Pavisys: A computer vision system for the inspection of locomotive pantographs," *Proc. of Pantograph Catenary Interaction Framework for Intelligent Control, Amiens*, 2011.

[6] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Robotics-DL tentative*. International Society for Optics and Photonics, 1992, pp. 586–606.

[7] D. A. Forsyth and J. Ponce, "A modern approach," *Computer Vision: A Modern Approach*, pp. 88–101, 2003.

[8] A. Ansar and K. Daniilidis, "Linear pose estimation from points or lines," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 5, pp. 578–589, 2003.

[9] D. Oberkampf, D. F. DeMenthon, and L. S. Davis, "Iterative pose estimation using coplanar points," in *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*. IEEE, 1993, pp. 626–627.

[10] R. M. Haralick, H. Joo, C.-N. Lee, X. Zhuang, V. G. Vaidya, and M. B. Kim, "Pose estimation from corresponding point data," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 19, no. 6, pp. 1426–1446, 1989.

[11] S. Hinterstoisser, S. Holzer, C. Cagniart, S. Ilic, K. Konolige, N. Navab, and V. Lepetit, "Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 858–865.

[12] M. Woo, J. Neider, T. Davis, and D. Shreiner, *OpenGL programming guide: the official guide to learning OpenGL, version 1.2*. Addison-Wesley Longman Publishing Co., Inc., 1999.

[13] D. Pagliari and L. Pinto, "Calibration of kinect for xbox one and comparison between the two generations of microsoft sensors," *Sensors*, vol. 15, no. 11, pp. 27 569–27 589, 2015.