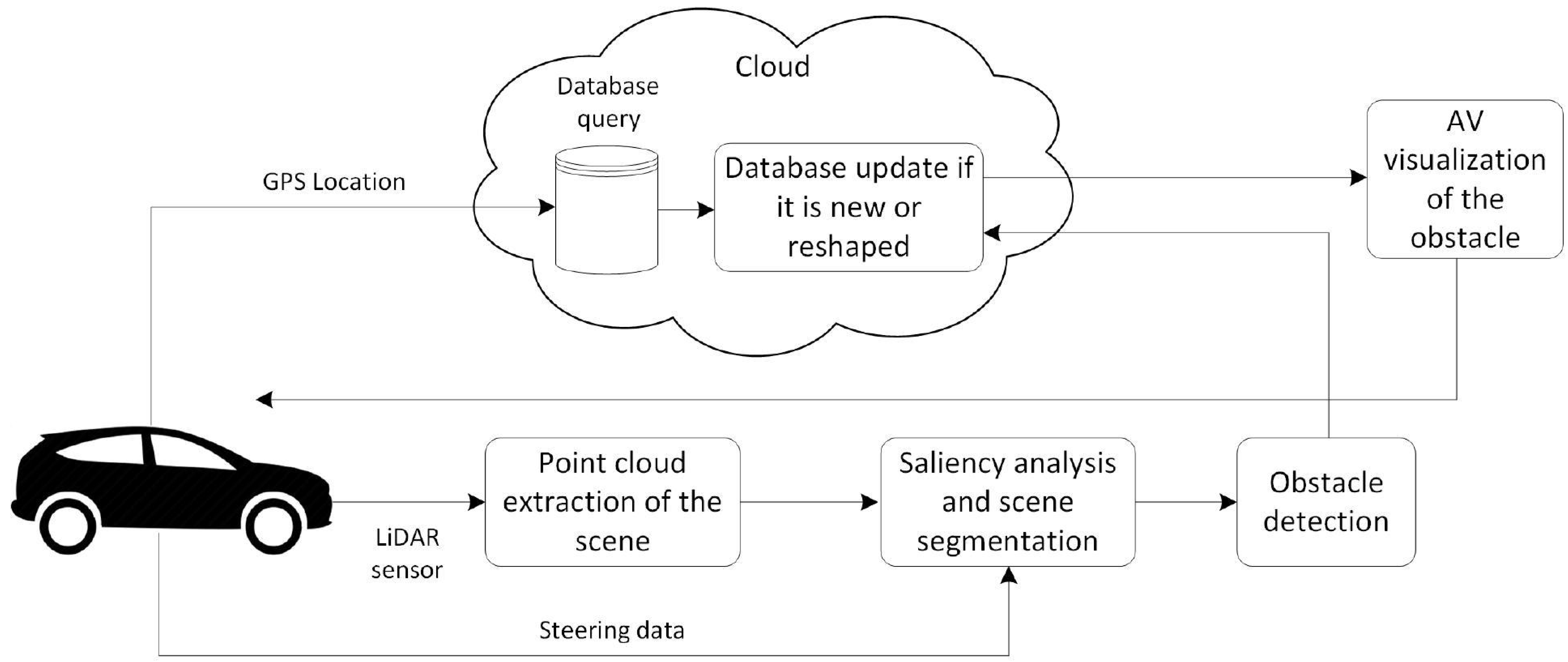

Fig. 1. Schematic diagram of the proposed methodology.

$\mathbf{V} = [\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots, \mathbf{v}_{m_i}] \in \mathbb{R}^{3\times m_i}$. Let's also denote with Ψ_j^k the set of the indices of the k nearest neighbors of point j. For a face f defined by three vertices $(\mathbf{v}_{j1}, \mathbf{v}_{j2}, \mathbf{v}_{j3})$, the outward unit face normal $\mathbf{n}_f$ is calculated by the following equation:

$$\mathbf{n}_f = \frac{(\mathbf{v}_{j2} - \mathbf{v}_{j1}) \times (\mathbf{v}_{j3} - \mathbf{v}_{j1})}{\|(\mathbf{v}_{j2} - \mathbf{v}_{j1}) \times (\mathbf{v}_{j3} - \mathbf{v}_{j1})\|} \quad (1)$$

The point normal $\mathbf{n}_j$, representing the normal of each point separately, is calculated as:

$$\mathbf{n}_j = \frac{\sum_{\forall \mathbf{n}_f \in \Psi_j^k} \mathbf{n}_f}{|\Psi_j^k|} \quad (2)$$

B. Saliency Map Estimation of the Point Cloud Scene

The purpose of this step is to calculate a metric of saliency for each vertex of a point cloud. Assuming point clouds without context information, saliency characterizes the geometric properties in a local neighborhood of points, i.e., high saliency values represent more perceptually prominent vertices which usually correspond to sharp corners (high-frequency spatial information). On the opposite, the geometrically least important points are those that lie in flat areas.

For the estimation of the saliency map, we implemented and modified the fusion technique presented in [33]. Instead of using guided normals of centroids, as in the original version [33], we now utilize normals for the points. This was performed to accelerate computations. Since the number of faces is usually approximately twice the number of vertices, the point normals are almost half the number of the centroid normals. For the sake of completeness, we present here our approach for the estimation of the saliency map of a point cloud scene, utilizing point normals.

Our fusion technique combines geometric saliency ($s^{(1)}$) with spectral saliency ($s^{(2)}$) features. The unique characteristics of each of these saliency features make the methodology more robust to point clouds acquired under real conditions, thereby being potentially affected by noise and outliers. The method processes each frame independently without examining past temporal information. Thus, as the methodology is applied for each point cloud in the sequence independently, for simplicity we omit the index i (indicating the frame number) from now on in the equations.

For a point cloud $\mathbf{P}$ with m vertices, a matrix $\mathbf{E} \in \mathbb{R}^{3m\times(k+1)}$ is constructed which includes in the first column the m point normals ($\mathbf{n}_j = [n_{j_x}, n_{j_y}, n_{j_z}]^T$) of each vertex j, $j = 1, \cdots, m$, respectively, and in the subsequent k columns the point normals of the k nearest neighbors of vertex j (i.e. $\mathbf{n}_{j\kappa} \in \Psi_j^k$). The salient features extracted by this approach capture global information since the matrix $\mathbf{E}$ is constructed using the point normals of the whole scene.

In order to exploit the geometrical coherence between neighboring normals, we apply Robust Principal Component Analysis (RPCA) to decompose the matrix $\mathbf{E}$ into a low-rank matrix $\mathbf{L} \in \mathbb{R}^{3m\times(k+1)}$ and a sparse matrix $\mathbf{S} \in \mathbb{R}^{3m\times(k+1)}$, as described in the appendix A. The matrix $\mathbf{L}$ consists of the low-rank values $\bar{\mathbf{n}}$ of the point normals $\mathbf{n}$, while the matrix $\mathbf{S}$ consists of the corresponding sparse values represented as $\dot{\mathbf{n}}$. The values of this matrix are zero (or to be more specific nearly zero) if the row (representing a neighboring patch of points) corresponds to point normals with very similar values, i.e., the vertex lies in a flat area, and very large values if the row corresponds to point normals with big dissimilarity (i.e., the vertex lies in a very sharp corner). The fact that most of the local patches Ψ_j^k of a 3D surface are piecewise flat confirms that the matrix $\mathbf{S}$ can be considered a sparse matrix.

$$\mathbf{S} = \begin{bmatrix} \dot{\mathbf{n}}_1 & \dot{\mathbf{n}}_{11} & \dot{\mathbf{n}}_{12} & \dots & \dot{\mathbf{n}}_{1k} \\ \dot{\mathbf{n}}_2 & \dot{\mathbf{n}}_{21} & \dot{\mathbf{n}}_{22} & \dots & \dot{\mathbf{n}}_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \dot{\mathbf{n}}_m & \dot{\mathbf{n}}_{m1} & \dot{\mathbf{n}}_{m2} & \dots & \dot{\mathbf{n}}_{mk} \end{bmatrix} \quad (3)$$

In other words, sparsity of the matrix is assumed because piecewise flat areas are the most dominant geometrical pattern in a 3D surface.

1) Estimation of the geometrical saliency (global approach): As the similarity of normals between neighboring