APPENDIX

TABLE I

EXAMPLE DATASET

| | $z_{i,1}$ | $z_{i,2}$ | $z_{i,3}$ | **Label** |
|---|---|---|---|---|
| $z_{1,j}$ | 0.20 | 0.21 | 0.3 | 2 |
| $z_{2,j}$ | 0.30 | 0.22 | 0.3 | 2 |
| $z_{3,j}$ | 0.30 | 0.16 | 0.55 | 0 |
| $z_{4,j}$ | 0.35 | 0.15 | 0.60 | 0 |
| $z_{5,j}$ | 0.40 | 0.14 | 0.65 | 0 |
| $z_{6,j}$ | 0.40 | 0.16 | 0.70 | 0 |
| $z_{7,j}$ | 0.95 | 0.55 | 0.50 | 1 |
| $z_{8,j}$ | 1 | 0.60 | 0.50 | 1 |
| $z_{9,j}$ | 0.35 | 0.18 | 0.55 | 0 |
| $z_{10,j}$ | 0.50 | 0.19 | 0.56 | 0 |
| $z_{11,j}$ | 0.25 | 0.21 | 0.72 | 0 |
| $z_{12,j}$ | 0.20 | 0.21 | 0.73 | 0 |
| $z_{13,j}$ | 0.30 | 0.22 | 0.74 | 0 |
| $z_{14,j}$ | 0.34 | 0.29 | 0.75 | 0 |
| $z_{15,j}$ | 0.15 | 0.26 | 0.76 | 0 |
| $z_{16,j}$ | 0.16 | 0.34 | 0.77 | 0 |
| $z_{17,j}$ | 0.01 | 0.01 | 0.1 | 3 |
| $z_{18,j}$ | 0.9 | 0.9 | 0.95 | 3 |

Legenda:
0= Inlier 1= Spatial Outlier 2= Temporal Outlier 3= Spatio-Temporal Outlier

*Lemma 1:* The construction of the lower approximation $\underline{B}(O)$ or the upper approximation $\overline{B}(O)$ of an *n–Outlier Set* $O$ converges if it exists an index $k$ such that the threshold does not vary anymore, i.e.

$$if \quad \overline{\tau}_k = \tau_k \quad then \quad \underline{B}_k(O) = \overline{B}_k(O) \tag{1}$$

*Lemma 1*

*Proof:*

Since $\underline{B}_k(O) = \{G_j \subseteq U : \overline{\omega}_{G_j} > \tau_i\}$ and $\overline{B}_k(O) = \{G_j \subseteq U : \overline{\omega}_{G_j} > \overline{\tau}_i\}$, we would to prove that:

$$\underline{B}_k(O) = \overline{B}_k(O) \ iff \ 1) \quad \underline{B}_k(O) \subseteq \overline{B}_k(O) \quad and \quad 2) \quad \overline{B}_k(O) \subseteq \underline{B}_k(O) \tag{2}$$

By definition

$$\forall G_j \subseteq \underline{B}_k(O) : \quad \overline{\omega}_{G_j} > \tau_k$$

and by hypothesis, $\exists \ \ k : \quad \overline{\tau}_k = \tau_k$, then:

$$\forall G_j \subseteq \underline{B}_k(O): \quad \overline{\omega}_{G_j} > \overline{\tau}_k \Longrightarrow G_j \subseteq \overline{B}_k(O)$$

Thus $\underline{B}_k(O) \subseteq \overline{B}_k(O)$.

Similarly, by definition,

$$\forall G_j \subseteq \overline{B}_k(O): \quad \overline{\omega}_{G_j} > \overline{\tau}_k$$

and by hypothesis, $\exists \;\; k: \quad \overline{\tau}_k = \tau_k$, then:

$$\forall G_j \subseteq \overline{B}_k(O): \quad \overline{\omega}_{G_j} > \tau_k \Longrightarrow G_j \subseteq \underline{B}_k(O)$$

and thus $\overline{B}_k(O) \subseteq \underline{B}_k(O)$. ■

*Proposition 1:* The measure computed in $K$ is an upper bound of the measure computed in $U$ such that:

$$d_p(U) \leq d_p(K), \quad \forall p \in U$$

$where\ d_p(U) = \sum_{j=1}^{k} d(p, N(p, p_j))\ and\ \ N(p, p_j)\ is\ the\ j\text{-}th\ nearest\ neighbor$ of $p$.

*Proof:* Let be $N(p, p_j)$ the $j\text{-}th\ nearest\ neighbor$ of $p\ in\ U$.

Two cases should be highlighted:

**Case 1**:

If $N(p, p_j) \in U\ and\ N(p, p_j) \in K,\ \forall j = 1, ..., k$ then $d_p(U) = d_p(K)$.

**Case 2**:

It exists an index $i$ such that $N(p, p_i) \in U\ and\ N(p, p_i) \notin K$.

In this case

$d_p(U) = \sum_{j=1, j \neq i}^{k} d(p, N(p, p_j)) + d(p, N(p, p_i))$

$d_p(K) = \sum_{j=1, j \neq i}^{k} d(p, N(p, p_j)) + d(p, N(p, \overline{p}))$

where $\overline{p} \in K$ is the *k–th nearest neighbor* of $p$ in $K$ to keep into account since $p_i \notin K$.

As $p_i$ is one of *k–nearest neighbors* of $p$ in $U$, the following inequality holds:

$d(p, N(p, p_i)) < d(p, N(p, \overline{p}))$ or, equivalently, $d_p(U) < d_p(K)$ ■

*Proposition 2:* The Outlier Set $O_K$, computed starting from Kernel Set $K$ is a superset of $O$ computed from $U$:

$$O_K \supseteq O$$

*Proof:* Let $O$ be the *n–Outlier Set* computed from $U$:

$$O = \{p_1, ..., p_n, p_{n+1}, ..., p_{\overline{n}} \in U \ / \ d_{p_1}(U) \geq ... \geq d_{p_{n(U)}} = d_{p_{n+1}}(U).. = d_{p_{\overline{n}}}(U) > d_{p_j}(U) \quad \forall j = \overline{n}+1, ..., N\}$$

where $d_{p_i}(U) = \sum_{j=1}^{k} d(p_i, N(p_i, p_j)), \quad \forall i = 1, ..., N$, is defined and computed on U.
We want to prove that:

$$\{p_1, ..., p_n, p_{n+1}, ..., p_{\overline{n}}\} \in O \ implies \ that \ \{p_1, ..., p_n, p_{n+1}, ..., p_{\overline{n}}\} \in O_k \tag{3}$$

By Proposition 1, the following inequality holds:

$$d_p(U) \leq d_p(K), \quad \forall p \in U$$

and in particular:

$$d_{p_i}(U) \leq d_{p_i}(K), \quad \forall i = 1, ..., \overline{n}$$

By definition of $n$-$Outlier\ Set$

$$d_{p_i}(U) \geq \ \tau, \ \forall i = 1, ..., \overline{n}$$

Thus:
$d_{p_i}(K) \geq \ \tau \ \forall i = 1, ..., \overline{n}$ implies $\{p_1, ..., p_n, p_{n+1}, ..., p_{\overline{n}}\} \in O_k$,
letting the thesis to hold. ■