FIG 2018 ISTANBUL

XXVI FIG Congress 2018

6–11 May 2018

ISTANBUL

ORGANISED BY

MAIN SUPPORTERS

T.C. ÇEVRE VE ŞEHİRCİLİK BAKANLIĞI

Tapu ve Kadastro

PLATINUM SPONSORS

esri THE SCIENCE OF WHERE

Leica Geosystems

Trimble

Building Change Detection using Semantic Segmentation on Analogue Aerial Photos

Evangelos MALTEZOS, Charalabos IOANNIDIS, Anastasios DOULAMIS and Nikolaos DOULAMIS

Laboratory of Photogrammetry, School of Rural and Surveying Engineering
National Technical University of Athens
Athens, Greece

INTRODUCTION – Change detection

Spatial-temporal urban monitoring in large-scale is critical in various engineering and civilian applications:

- Identification of informal settlements and constructions
- Disaster management
- Population estimation
- Urban and rural planning
- Transportation
- Augmented reality
- 3D city modeling
- Housing value
- Updating of 3D cadastral objects and databases

AUTOMATIC CHANGE DETECTION

The wider scope of the automatic change detection is a research topic in over two decades. Approaches:

1. Change enhancement: Indicates only the position and the magnitude of the change
2. From-to: Indicates not only the position and the magnitude of the change but also its nature (e.g., **building changes**, vegetation changes, etc.)

Building change detection includes several challenges and difficulties:

- Radiometric and atmospheric correction and calibration
- Geometric correction and data registration
- Multi-modal data fusion
- Inherent artifacts of the used data
- Differences in viewpoint and surrounding environment
- Complex shape and size of the building structures
- Use of efficient algorithms with generalization properties

BUILDING CHANGE DETECTION TECHNIQUES

Building change detection techniques dependent on the data source employed:

1. **Use only airborne or satellite imagery data**: fully exploit image information, additional information/indices, band ratios, etc.
   (-) sensitive on shadows and urban objects with similar pixel values, e.g., building rooftops vs. roads

2. **Use only 3D information**: implement direct 3D point cloud processing either derived from Dense Image Matching (DIM) techniques or by LIDAR data or a combination of the two
   (-) incorrect assignments of building changes when vegetation is significantly grows up between two time periods

3. **Use a combination of both of data sources**: Fusion of image and 3D information
   (+) appropriate for detecting 2D/3D building changes

Requirements

i) Co-registration of the data coming from different sources
ii) Additional acquisition and processing cost of the multi-modal information
iii) LIDAR data or additional information are not always available

DENSE IMAGE MATCHING

- **Dense Image Matching (DIM):** High quality dense 3D point cloud of an area (Digital Surface Model-DSM) calculating the depth information by stereo matching (disparity estimation)
- The last years more and more researchers use image processing techniques to automatically extract 3D information using multiple 2D imagery that cover an area of interest

(+) Cost effectiveness, flexibility, exploitation of available old aerial photos

- Typical problems during DIM:
  - Texture-less areas
  - Repetitive patterns
  - Complex scene
  - Radiometric differences (e.g., shadows)
  - Moving objects (e.g., cars)
  - Occlusions
  - Low quality of the images (for case of scanned analogue photos)
  - Geometry of the availiable stereo-pairs

Impact:

- Surface roughness
- Local deformations at the boundaries of the objects
- Excessive interpolations
- Radiometric deformations to the corresponding orthoimages

AIM OF THE STUDY

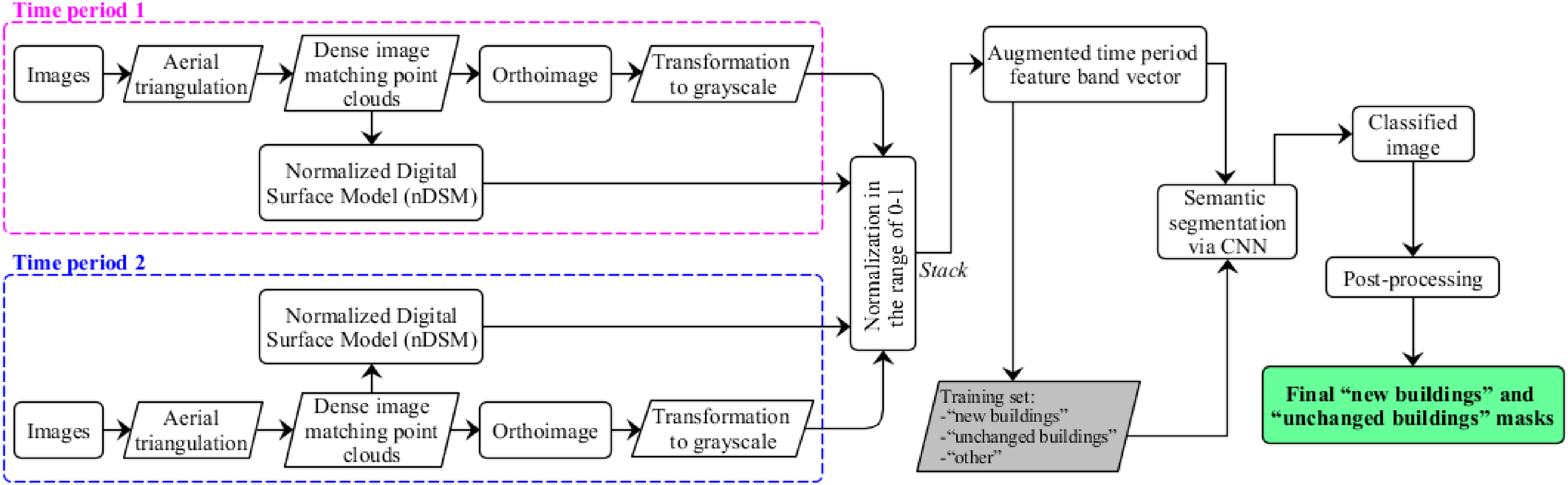
A procedure of **one-shot building change detection framework** through **semantic segmentation** (deep CNN) using scanned **analogue aerial photos**

Semantic segmentation is a pixel-wise classification of images by implementing a **deep neural network** scheme such as **Convolutional Neural Networks** (CNNs) under a **supervised setting**

Contribution:

- ✓ Full exploitation of one type of data with i) poor and/or limited spectral capabilities, ii) medium quality and iii) no additional information
- ✓ Fusion of scanned analogue aerial photos with modern derived remote sensing data, e.g. UAV
- ✓ Fusion of image (orthoimages) and 3D information (DIM/nDSMs).
- ✓ Use of sophisticated deep learning model (Convolutional Neural Network - CNN)

Workflow



An augmented time period feature band vector is created by fusing 3D geospatial information: orthoimage and 3D point cloud

A small training set for the classes of "new buildings", "unchanged buildings" and "other" is created from the same dataset

The training set and the augmented time period feature band vector are fed as input into a CNN

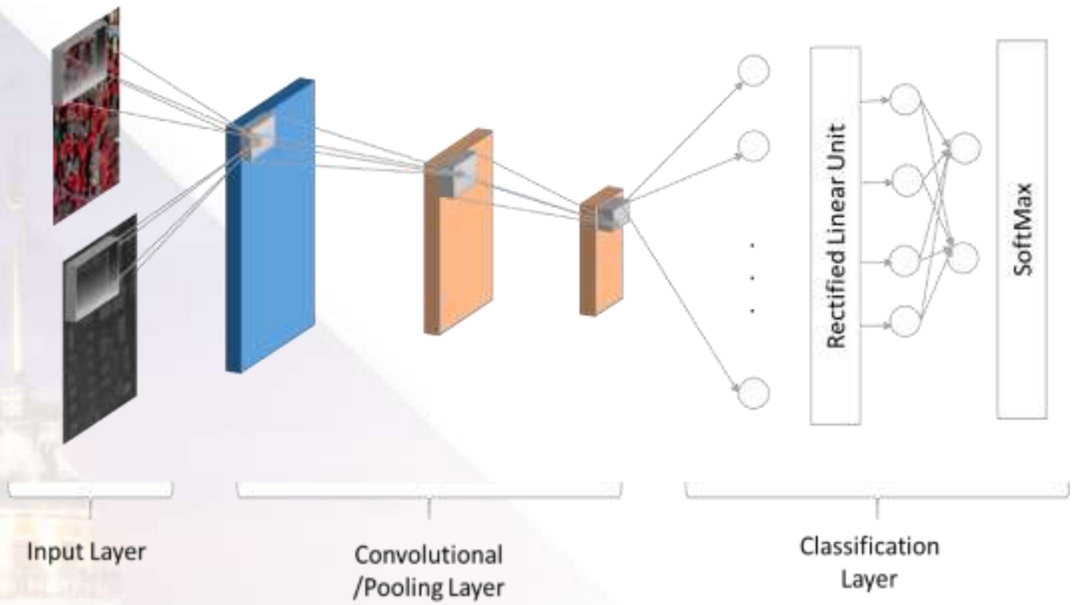
The "new buildings" and the "unchanged buildings" masks are then processed to eliminate noise taking into account the spatial coherency properties

The Convolutional Neural Network (CNN) structure

A CNN classifier has two main components:

- *Convolutional layer*: filters the input data to maximize the classification accuracy
  It is a feature extractor that employs convolution filters (transformations) to the input data → Optimization of:
  the classification performance + spatial coherency (it is an important property of a deep CNN model since spatial characteristics significantly affect building change detection accuracy)

- *Classification layer:* supervised learning model with a capability of transforming the inputs from the convolutional layer to the labeled classes in pixel level



Example of a deep CNN

Implementation of CNN:

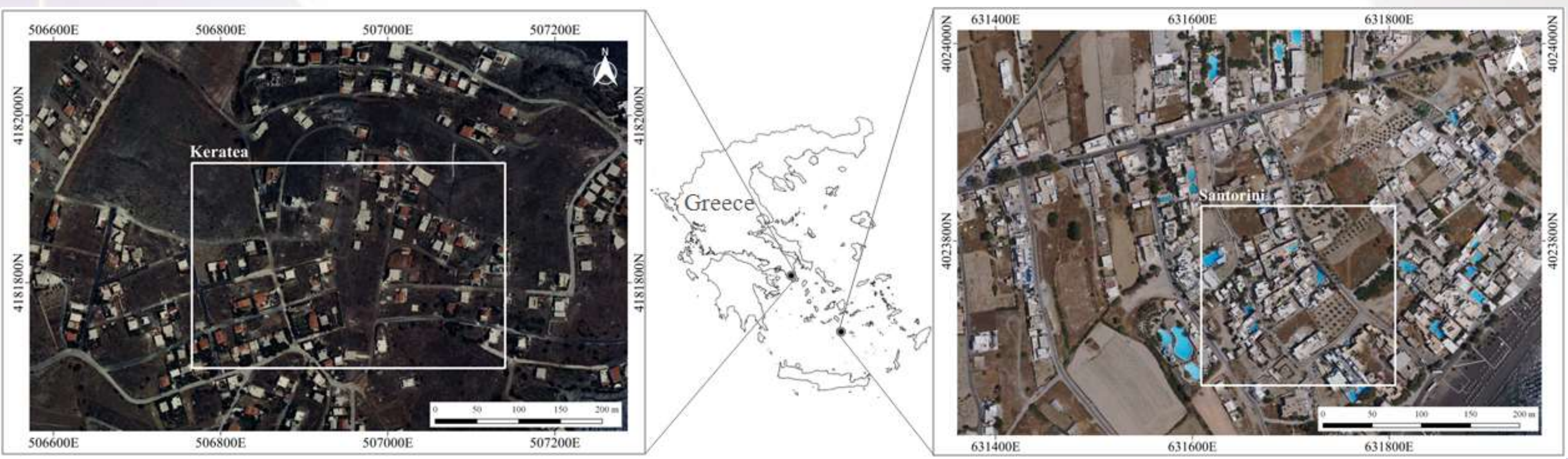
Input layer: AUFEV (Patch size: 5x5)
1st layer: Convolutional layer ($C1=3\times d$ trainable filters of dimension 3×3),
2nd layer: Convolutional layer ($C2=3\times C1$ trainable filters of dimension 3×3),
Classification layer: MultiLayer Perceptron (MLP) classifier

In this study: The final output of the CNN is a classified image of classes i) new buildings, ii) unchanged buildings and iii) other

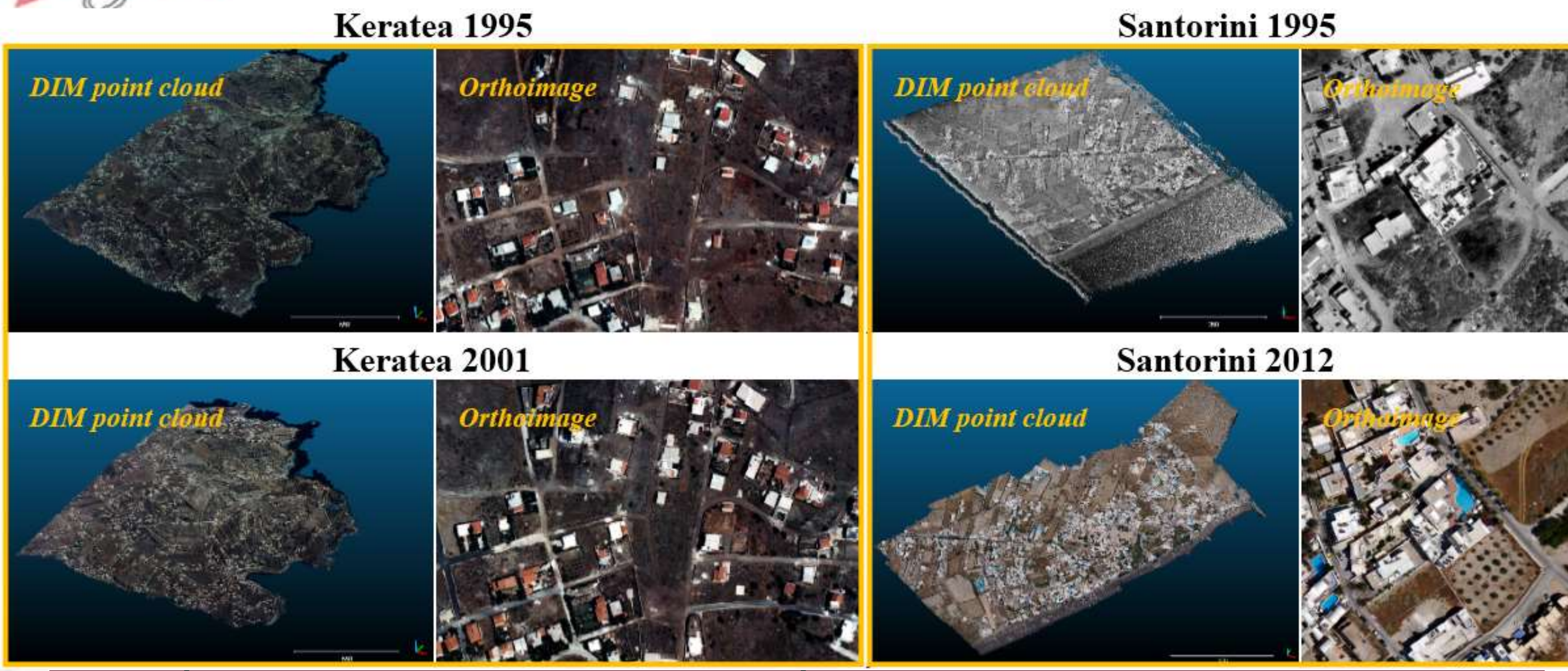
EXPERIMENTAL RESULTS – Study areas

<u>Keratea</u> → Buildings: small detached houses with sloped or flat roof surfaces
Vegetation: sporadic, often high

<u>Santorini</u> → Buildings: very complex + objects on the rooftops
Vegetation: moderate



EXPERIMENTAL RESULTS – Datasets



Study area	Time Period 1			Time Period 2		
	Type of images	nDSM	Orthoimage	Type of images	nDSM	Orthoimage
Keratea	RGB Scanned analogue photos	from DIM GSD=21 cm	GSD=21 cm	RGB Scanned analogue photos	from DIM GSD=21 cm	GSD=21 cm
Santorini	Grayscale Scanned analogue photos	from DIM GSD=7 cm	GSD=7 cm	RGB UAV digital images	from DIM GSD=7 cm	GSD=7 cm

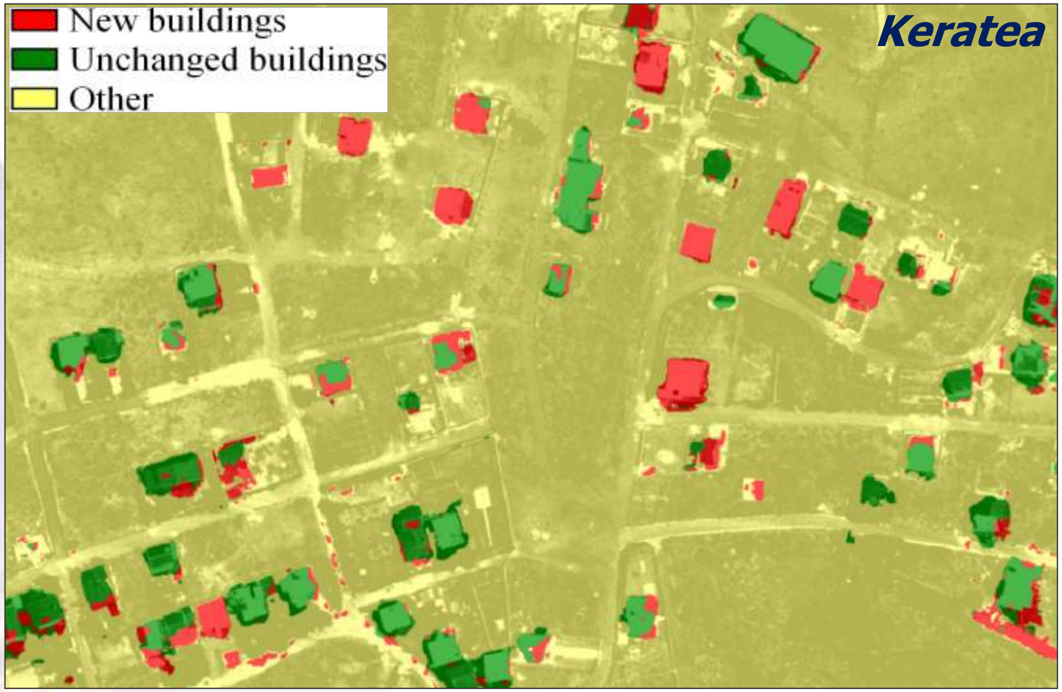
EXPERIMENTAL RESULTS – Learning set

Three classes are used for the semantic segmentation (training set=80%, validation set=20%):

- **New buildings**: buildings that were constructed at the second time period
- **Unchanged buildings**: buildings that remain unchanged at both time periods
- **Other**: ground, high and low vegetation, cars, insignificant structures (either they have undergone changes over the time or not)

Study area	Class: New buildings	Class: Unchanged buildings	Class: Other	Percentage sum
	Percentage on the block tile	Percentage on the block tile	Percentage on the block tile	
Keratea	0.04%	0.07%	0.21%	**0.33%**
Santorini	0.30%	0.25%	0.79%	**1.35%**

EXPERIMENTAL RESULTS – Semantic segmentation via CNN



Computational time

Keratea → Learning time: 3 min, Test time: 30 sec
Santorini → Learning time: 15 min, Test time: 3 min

Post process:

Step 1 → Extraction of the "new buildings" and "unchanged buildings" masks
Step 2 → 7×7 morphological median operator (twice)
Object area thresholds: Keratea: 20 m^2 , Santorini: 4 m^2

EVALUATION AND COMPARISONS

- **Evaluation metrics:** Completeness (C_m), Correctness (C_r) and Quality (Q)

$$C_m\,(\%) = \frac{\|TP\|}{\|TP\|+\|FN\|}\cdot 100;\quad C_r\,(\%) = \frac{\|TP\|}{\|TP\|+\|FP\|}\cdot 100;\quad Q\,(\%) = \frac{\|TP\|}{\|TP\|+\|FP\|+\|FN\|}\cdot 100$$

True Positives (TP): correct assignments
False Positives (FP): false assignments
False Negatives (FN): missing assignments

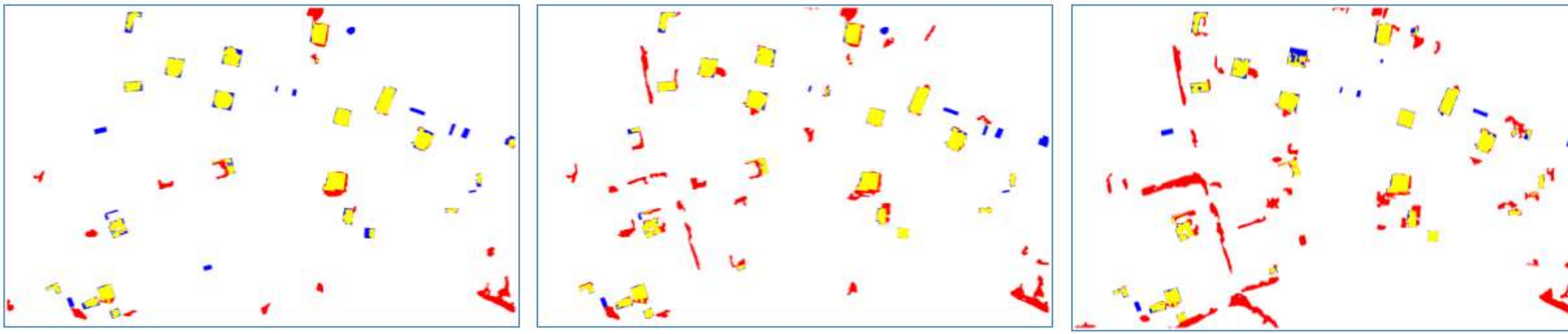
- **Reference data:** Manually digitized from the orthoimages and the DIM/DSM

- **Comparison with shalow machine learning:**
  - Support Vector Machine (SVM) with Linear kernel
  - Support Vector Machine (SVM) with Radial Base Function (RBF) kernel

EVALUATION AND COMPARISONS

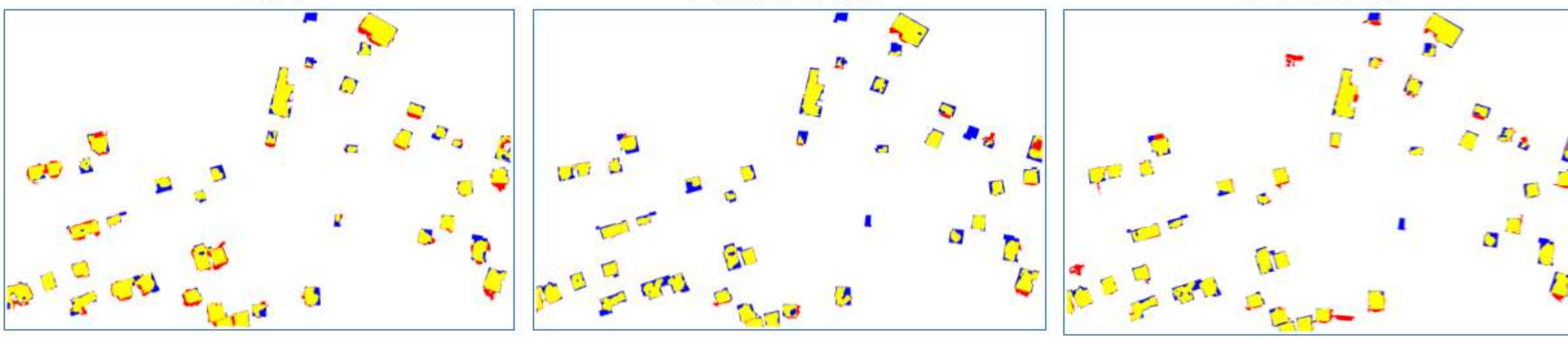
Study area: KERATEA

CNN | Linear SVM | RBF SVM



Pixel-wise results of class **"new buildings"** TP FP FN

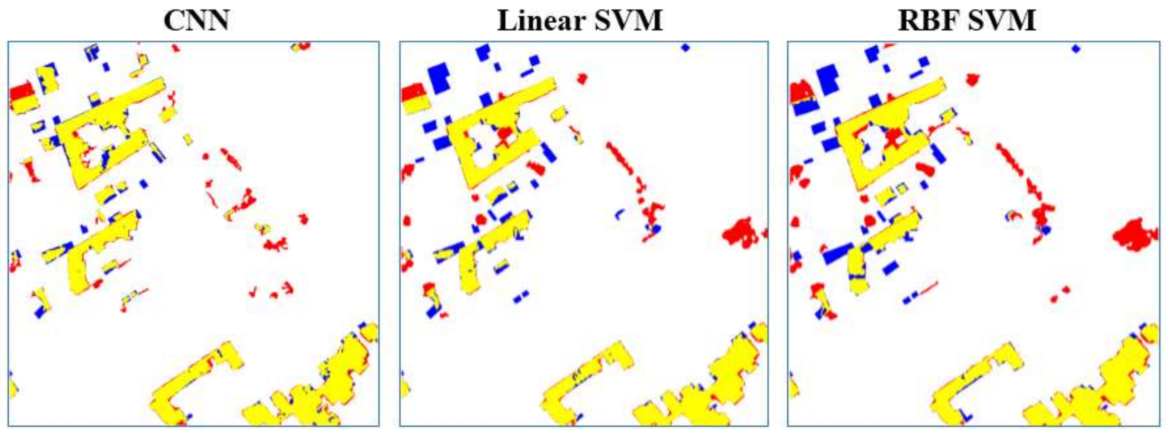
CNN | Linear SVM | RBF SVM



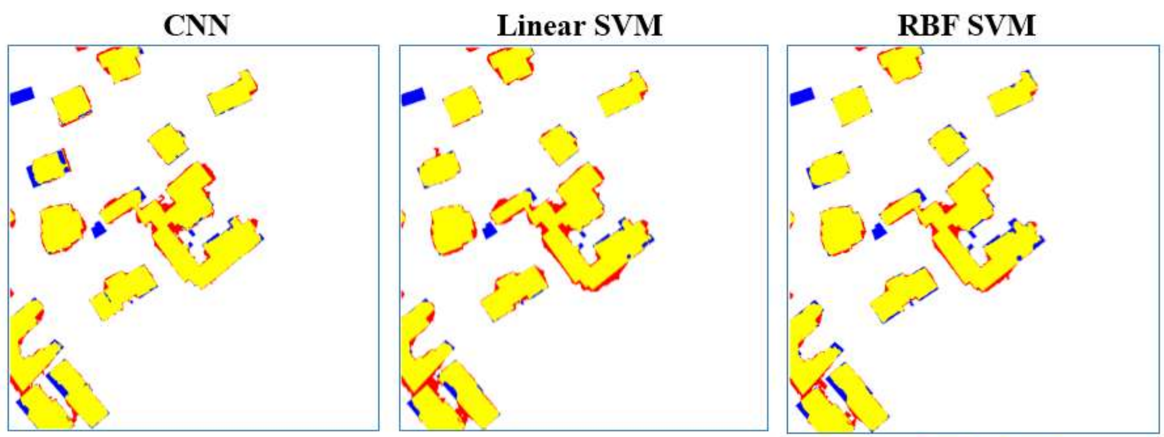
Pixel-wise results of class **"unchanged buildings"** TP FP FN

EVALUATION AND COMPARISONS

Study area: Santorini



Pixel-wise results of class “**new buildings**” TP FP FN



Pixel-wise results of class “**unchanged buildings**” TP FP FN

EVALUATION AND COMPARISONS

FP of the "new buildings":

- Pixels that assigned as "new buildings" while should be assigned as "unchanged buildings"
- Artifacts coming from occlusions or mismatches during the DIM at one time period expressed as excessive interpolations during the rasterization of the nDSMs
- Insignificant objects (e.g. umbrella arrays) that present similar characteristics (height and image pixel values) with building changes.

FN of the "new buildings":

- Insufficient 3D reconstruction of a building during the DIM of the second time period or by lack of representative training samples

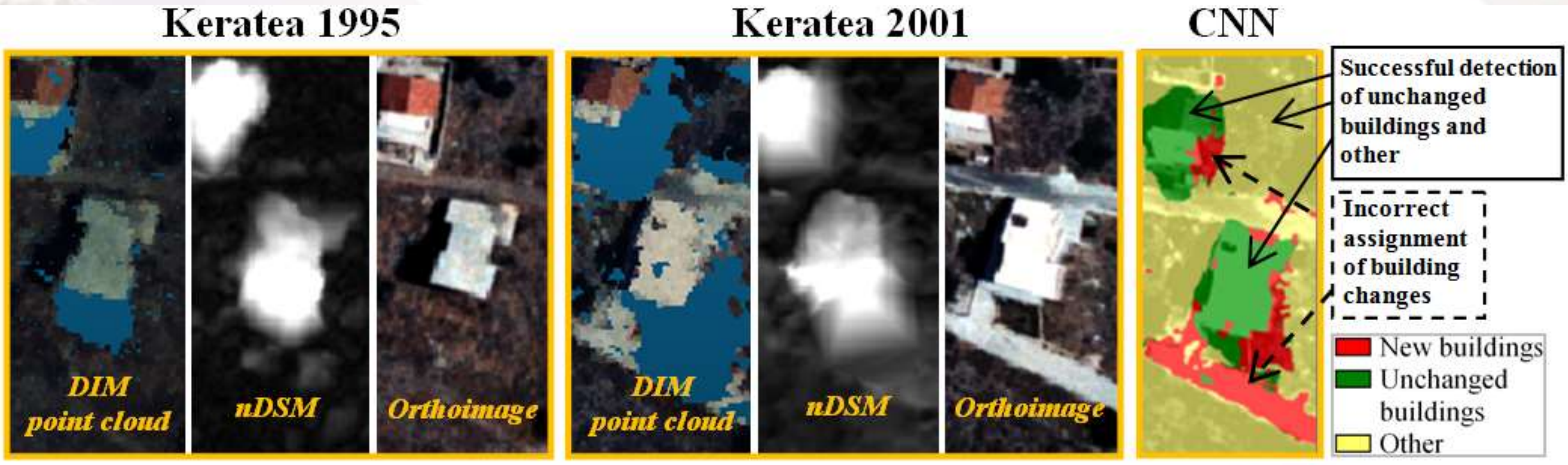
EVALUATION AND COMPARISONS

FP of the "unchanged buildings":

- Excessive interpolations during DIM
- Remaining misregistration errors between the datasets of both time periods

FN of the "unchanged buildings":

- Mismatches during the DIM (i.e., insufficient 3D reconstruction of a building at one or both time periods)
- Incorrect detection of the ground points during the calculation of the nDSM (e.g., incorrect assigment of buildings as ground at one or both time periods)



EVALUATION AND COMPARISONS

Study area	Classes	CNN			Linear SVM			RBF SVM		
		C_m (%)	C_r (%)	Q (%)	C_m (%)	C_r (%)	Q (%)	C_m (%)	C_r (%)	Q (%)
Keratea	New buildings	**75.0**	**67.8**	**55.3**	83.4	53.9	48.7	79.0	43.00	38.6
Keratea	Unchanged buildings	**81.9**	**83.2**	**70.3**	68.2	91.8	64.3	79.3	86.6	70.7
Santorini	New buildings	**87.0**	**83.0**	**73.8**	79.4	78.7	65.3	76.1	72.5	59.0
Santorini	Unchanged buildings	**92.5**	**89.7**	**83.7**	94.3	87.0	82.6	92.1	90.8	84.2
	Average New buildings	**81.0**	**75.4**	**64.6**	81.4	66.3	57.0	77.6	57.8	48.8
	Average Unchanged buildings	**87.2**	**86.5**	**77.0**	81.3	89.4	73.5	85.7	88.7	77.5

- Total computational time of SVMs → less than 1 min
- The proposed CNN provides balanced rates and better classification performance compared to SVM classifiers

CONCLUSIONS

- ✓ The CNN results show the applicability and functionality of deep learning models to detect building changes in terms of various kinds of buildings structures, pixel resolutions and types of data
- ✓ The use of a better CPU and more RAM capacity can significantly speed the processing
- ✓ To overcome misclassification problems:
  - More representative training samples
  - Building reconstruction techniques
  - Additional morphological criteria
- ✓ Reliable building change detection maps for an initial spatial-temporal analysis and predictive assessment
- ✓ Future work is needed to explore and evaluate the performance of several deep learning schemes for building change detection