Thesis title: **Mapping between Views for Multi-View 3D Class Models**

Contact persons: Prof. Dr. Peter Eisert, (eisert@informatik.hu-berlin.de)
Dr. Ismail El Sayad (ismail.el.sayad.@hhi-extern.fraunhofer.de)


Thesis description

Background:

Imagine a busy street in downtown Rome or Beijing. There might be cars coming from all possible directions, pedestrians walking on the sidewalk, or scooters crossing the street. The ability to interpret the scene, recognize the *buildings from different views* within, estimate their locations and poses is crucial [1]. Recently, a number of methods have brought new ideas into the problem of representing object categories from multiple views by using a rough 3D model, on top of which the typical distribution of appearance is learned [2]. These methods have the advantage of yielding a compact view-point invariant representation of the object category, as opposed but fail to accommodate *intra-class 3D* appearance variability [1-2].

An alternative strategy is to use *multiple views* of *multiple instances* to construct a model of the distribution of 3-dimensional shapes of the object class. Such a model would allow recognition of many entirely novel views. This is an open problem, which is still unsolved [3].

Problem statement:

We work on constructing a 3D class model for multi-view Building class Detection using visual language modeling [4] for high-level feature extraction. In addition, machine learning techniques will be used for model parameters estimation [5]. One of the parameters of the 3D class model is the **k-Views mapping**. This parameter describes dependencies between high-level features of images in two different views within the same class.

The work of the master thesis includes but is not limited to the following items:

1. Review existing algorithms in **point matching techniques** such as shape context [6] and select those most applicable for 3D class model.
2. Implementation and experimental comparison of the selected algorithms. Criteria for the comparison will include computational complexity, and accuracy.
3. Designing a new algorithm (an improvement of an existing method or a combination of existing methods) that can give a best performance under the context stated above.

References:

[1] Sun, M., Su, H., Savarese, S., & Fei-Fei, L. (2009). A multi-view probabilistic model for 3d object classes. *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 1247–1254).

[2] Saxena, A., Sun, M., & Ng, A. Y. (2009). Make3D: learning 3D scene structure from a single still image. *IEEE transactions on pattern analysis and machine intelligence*, *31*(5), 824-40. doi:10.1109/TPAMI.2008.132

[3] Chiu, H. P., Kaelbling, L. P., & Lozano-Pérez, T. (2007). Virtual training for multi-view object class recognition. *IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07* (pp. 1–8), *2007.*

[4] El Sayad, I., Martinet, J., Urruty, T., Benabbas, Y., & Djeraba, C. (2011). A semantically significant visual representation for social image retrieval. *IEEE International Conference on Multimedia and Expo (ICME), 2011* (pp. 1–6).

[5] El Sayad, I., Martinet, J., Urruty, T., & Djeraba, C. (2010). Toward a higher-level visual representation for content-based image retrieval. *Multimedia Tools and Applications*. doi:10.1007/s11042-010-0596-x

[6] Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(4), 509-522. doi:10.1109/34.993558