A Novel Visual Organization Based on Topological Perception

Yongzhen Huang[1], Kaiqi Huang[1], Tieniu Tan[1], and Dacheng Tao[2]

[1] National Laboratory of Pattern Recognition, Institute of Automation
Chinese Academy of Sciences, Beijing, China
{yzhuang, kqhuang, tnt}@nlpr.ia.ac.cn
[2] Cognitive Computing Group, School of Computer Engineering
Nanyang Technological University, Singapore
{dacheng.tao}@ieee.org

Abstract. What are the primitives of visual perception? The early feature-analysis theory insists on it being a local-to-global process which has acted as the foundation of most computer vision applications for the past 30 years. The early holistic registration theory, however, considers it as a global-to-local process, of which Chen's theory of topological perceptual organization (TPO) has been strongly supported by psychological and physiological proofs. In this paper, inspired by Chen's theory, we propose a novel visual organization, termed computational topological perceptual organization (CTPO), which pioneers the early holistic registration in computational vision. Empirical studies on synthetic datasets prove that CTPO is invariant to global transformation such as translation, scaling, rotation and insensitive to topological deformation. We also extend it to other applications by integrating it with local features. Experiments show that our algorithm achieves competitive performance compared with some popular algorithms.

1 Introduction

Thirty years ago, Marr's primal sketch theory [1], [2] claimed that the primitives of visual information are simple components of forms and their local geometric properties, e.g., line segments with slopes. Influenced by this famous theory, the description of visual information has made great progress in many computer vision applications, e.g., various image descriptors have been developed recently. One of the representatives is the Scale-Invariant Feature Transform (SIFT) [3]. Comparative studies and performance evaluation on image descriptors can be found in [4] wherein SIFT based descriptors achieve the best performance. SIFT descriptor constructs a histogram by accumulating the weighted gradient information around an interest point, as well as orientation reassignment around dominant directions. This strategy enhances the insensitiveness to noises and robustness to local rotation. However, it ignores the global information which is essential to encode the deformation invariability. In addition, it encounters a great challenging to describe a meaningful structure. Variances in position and orientation are often substantial and suggest that a scale grouping rule is insufficient to achieve appropriate association of image fragments [5]. Although improvements emerge in specific applications, how to effectively organize local features is still very difficult and largely unexplored in computer vision. Perhaps, it is necessary to reconsider the intrinsic of the problem: what are the primitives of visual perception?

In this sense, one may need, like Gestalt, the whole as a guide to sum the parts. Chen's theory of topological perceptual organization (TPO) [6], [7], which assumes that wholes are coded prior to analysis of their separable properties or parts, is a view inherited from the Gestalt concept of perceptual organization. TPO is superior to the early feature-analysis theory in organizing local features and describing topological structures. Details are described in the next section.

In this paper, we propose a novel visual organization based on Chen's theory, termed computational topological perceptual organization (CTPO). To prove the effectiveness of the proposed CTPO, we conduct empirical justifications on synthetic datasets. Inheriting from the superiority of Chen's theory, CTPO is invariant to translation, scaling, rotation and insensitive to topological deformation. Besides, we integrating CTPO with popular local features and extend it for object categorization. Experiments show that CTPO achieves competitive performance compared with top level algorithms.

2 Topological Perceptual Organization

Chen's topological perceptual organization theory [6], [7] is a view inherited from the early holistic registration theory and the Gestalt psychology. Furthermore, it developed the Gestalt psychology and used thorough experiments to well support that 1) visual perception is from global to local; 2) wholes are coded prior to local; 3) global properties can be represented by the topologically invariant features.

The most important concept in Chen's theory is "perceptual object", which is defined as the invariant in topological transformations. A topological transformation [7] is, in mathematical terminology, a one-to-one and continuous transformation. Intuitively, it can be imagined as an arbitrary "rubber-sheet" distortion, in which neither breaks nor fusions can happen, e.g., a disc smoothly changing to a solid ellipse. Klein's Erlangen Program [8] shows that the topological transformation is the most stable among all geometric transformations. Moreover, it has been proved by neuroscience research [9] that, in human vision system, the topological transformation is the strongest stimulation among all the transformations.

Chen's theory has been strongly supported by psychological and biological experiments. For example, they tested the response of bees on different topological shapes. In the first stage, bees are trained to find a specific object (a ring) with a reward of sugar water. In the second stage, the reward is removed to test bees' reaction on different new object. In test 1, results show that bees are in favor of the diamond image which is topologically identical to the ring image. In test 2, results show that bees have no marked feeling for both shapes because the ring and the square share the same topological structure although they are different in local features. Apparently, experiments show that it is hard to differentiate topologically identical shapes. Experimental results are consistent with Chen's assumption that the topological structure is the fundamental component of the visual vocabulary. A comprehensive description and discussion of his theory may be found in a whole issue of "Visual Cognition" [7], [10], [11]. Chen's theory has opened new lines of research that are worthy of attention from not only visual psychologists but researchers in computer vision.

3 Computational Topological Perceptual Organization

Although Chen's theory of topological perceptual organization makes great progress in visual psychology, it does not provide a mathematical form to describe the topological properties of a structure. In this paper, we propose a computational topological perceptual organization (CTPO) to bridge the gap between Chen's psychological theory and the computer vision theory.

It is worth emphasizing that the concept of "global" in Chen's theory does not refer to the visual information of a whole image but a topological structure of an object. For a whole complex image, it is hard to describe its topological properties by Chen's theory currently.

3.1 Topology space

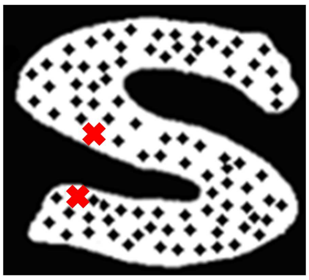


Fig. 1. An example of illustrating the difference between Euclidean distance and geodesic distance.

The core idea of topological perceptual organization is that "perceptual object" preserves invariants in topological transformation. According to Chen's theory, the connectivity and the number of holes are essential to describe the properties of a topological structure because they are invariant measurements in topological transformations. We adopt the distance between pairs of two pixels to describe the topological structure. However, the Euclidean measure is apparently not a good candidate. For example, in Fig. 1, two red crosses appear deceptively close, measured by their spatial Euclidean distance. But their spatial connectivity distance is large and reflects their intrinsic spatial relationship. Intuitively, it is necessary to define the topological property in such a space, wherein the distance between two points can reflect the connectivity of a structure and a group of such distances can reflect the number of holes. Therefore, the geodesic distance or the shortest path is a good choice.

The geodesic distance, however, encounters a problem: how to define global properties (e.g., connectivity) in a discrete set of image pixels? Fortunately, the tolerance space [7] can be applied to deal with this problem.

Definition 1. *Let X be a finite set of discrete dots. The tolerance ξ refers to the range within which detailed variations are ignored for attaching importance to global properties. The set of dots X together with the tolerance is a tolerance space denoted as (X, ξ).*

The tolerance and the global properties of a discrete set, therefore, are mutually dependent concepts. Fig. 2 shows an example. For human vision system, the tolerance ξ means the shortest noticeable distance. Under this definition, two points are connective only if they are in a specific tolerance. The notion of tolerance space not only resolves the problem of describing connectivity in a discrete set of image pixels but also builds the relationship between the scale and the structure.

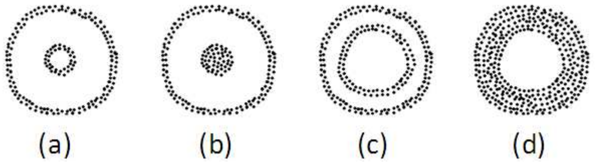


Fig. 2. Illustration of the tolerance space. If the tolerance is 1mm, (a) and (c) share an identical topological structure (both of them are "two rings"). This description matches our holistic perception. As the tolerance increases to a specific value, e.g., 4mm, (a) and (b) share an identical topological structure. So do (c) and (d).

Based on the above analysis, we propose a topology space d^*, defined as:

$$d^* = g(d'), \tag{1}$$

where g is the operation of calculating geodesic distance. The distance between two pixels in the space d' is computed as:

$$d'(i,j) = \begin{cases} d_1(i,j) + \lambda \times d_2(i,j), & \text{if } d_1(i,j) < \xi \\ \infty, & \text{otherwise} \end{cases}, \tag{2}$$

where d_1 and d_2 denote the spatial Euclidean distance and the intensity difference respectively, λ is a tradeoff parameter, and ξ is the tolerance in **Definition 1**.

The topology space considers the connectivity in two aspects: the spatial distance and the intensity difference. The latter one weakens the impact of pixels that are greatly different from their neighbors in terms of intensity.

3.2 Quotient distance histogram

In this section, we construct a quotient distance histogram (QDH) in the aforementioned topology space to describe topological properties. We adopt the quotient between $d^*(i,j)$ and $d'(i,j)$ as the vote to construct a histogram. This intuition is reasonable because $d^*(i,j)$ contains rich structural information and the quotient between $d^*(i,j)$ and $d'(i,j)$ is scale-invariant. The value of each bin in the histogram is given by:

$$h(k) = \sum_{i=1}^{n} \sum_{j=i+1}^{n} I\left(\theta(i,j) \in B(k)\right), \tag{3}$$

$$\theta(i,j) = d^*(i,j) \; / \; d'(i,j), \tag{4}$$

where n is the number of pixels in the structure, I is the indicator function, $B(k)$ is the range of k^{th} bin, $d^*(i,j)$ and $d'(i,j)$ are defined in Eq. (1) and Eq. (2). Note that $d'(i,j)$ may be infinite, and then $d^*(i,j)$ is infinite too. In this special case, we set $\theta(i,j)$ to infinity.

QDH can describe various topological structures. Its effectiveness is demonstrated in the experiment of Section 4.1.

It is important to understand that QDH is different from the Geodesic Intensity Histogram (GIH) [12] or the inner-distance [13] because: 1) QDH reflects the global statistical property by the quotient between $d^*(i,j)$ and $d'(i,j)$. The inner-distance is defined as the shortest path between mark points of a shape silhouette. Under this definition, even a "W" structure and a "S" structure cannot be differentiated by GIH. Besides, QDH is scale-invariant while GIH is not; 2) Besides the spatial distance, QDH also consider the intensity difference of image, thus it can be applied for gray images, while the inner-distance can only be used for binary images; 3) QDH is associated with the tolerance space, which emulates the visual characteristics of human vision system; and 4) the motivations of QDH and the inner-distance are totally different. QDH is inspired by a significant visual psychology theory.

4 Experiments

In the experimental part, we first conduct empirical justifications on synthetic datasets to differentiate basic topological structures. Second, we combine it with another model (EBIM [14]) and apply it for object categorization.

There are three parameters in CTPO: **1)**, ξ is the tolerance. To obtain robust performances, it is necessary to set a group of tolerances, e.g., $1, \sqrt{2}, \sqrt{3}, 2, \sqrt{5}$ in our experiments. **2)**, λ is a tradeoff parameter between spatial distance and intensity difference. In our experiments, we empirically set λ to 0.5, which is robust in most of cases. **3)**, $B(k)$ is the range of each bin in the CTPO histogram. It is insensitive to the performance. In fact, we calculate the probability distribution of $\theta(i,j)$ on the training set (or a part of original samples), and then equally divide the distribution. Then, every two division points can determine the range of a bin.

4.1 Structure classification

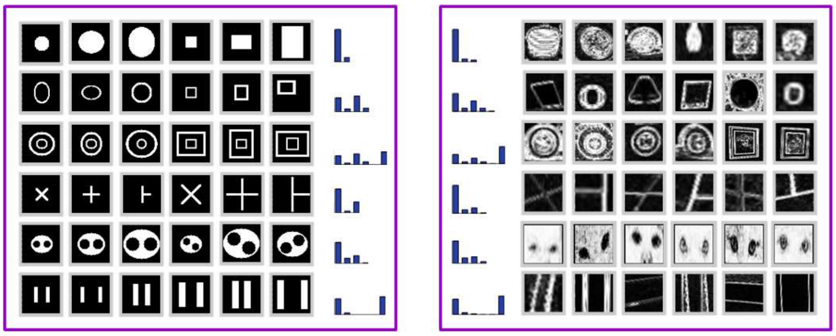
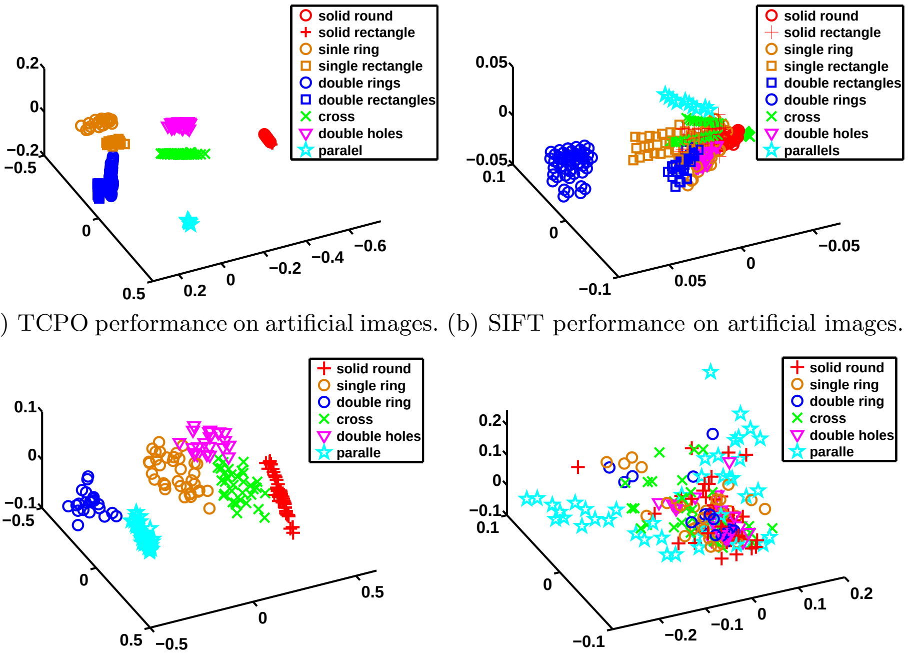


Fig. 3. Examples of artificial (left) and real (right) images. The histogram in each row is the mean quotient distance histogram corresponding to the images of the row.

In this experiment, we design two experiments to demonstrate the discriminative ability of CTPO on different topological structures tested in Chen's psychological research. In the first experiment, artificial images are used. For each category of topology structures, we first draw an initial image, and then produce some variants by using topological transformations defined in Section 2, e.g., translation, scaling rotation and deformation. In the second experiment, we choose real patches sampled from the images. Fig. 3 shows some examples of artificial shapes and real image patches (with noises). From top to bottom, they are the round, the single ring, the double rings, the cross, the double holes and the parallel. We also show their corresponding mean histograms in Fig. 3. Each mean histogram is the mean over all quotient distance histograms of a class of artificial shapes or real image patches. The mean histogram is almost equivalent for artificial shapes and real image patches in the same row, which proves that our algorithm can be applied to real images and robust to noises.

We compare the proposed quotient distance histogram with SIFT-like descriptor. To implement the SIFT-like descriptor, we use 8 orientations for the gradient histogram calculation in all divided (4×4) areas to construct a standard feature vector (128 dimensionality). Then the Isomap algorithm [15] is applied for dimensionality reduction to intuitively compare their ability of preserving the intrinsic properties of the topological structure. Fig. 4 shows the experimental



(a) TCPO performance on artificial images. (b) SIFT performance on artificial images.

(c) TCPO performance on real images. (d) SIFT performance on real images.

Fig. 4. (Please view in color) Comparison between CTPO and SIFT-like descriptor in preserving the geometric nature of the topological structures manifold. In (a) and (c), different topological structures are effectively differentiated by our approach for artificial and real images respectively. In (b), some different topological structures of artificial images are mixed by SIFT-like descriptor. In (d), SIFT-like descriptor has little ability of differentiating various topological structures.

results. Our approach (CTPO) greatly performs the SIFT-like descriptor in both artificial and real images. The experimental results are consistent with Chen's theory, and prove the effectiveness of the proposed topology space as well as the quotient distance histogram in differentiating various topological structures.

4.2 Object categorization

According to Chen's theory, the topological perceptual organization (TPO) should be applied to images or image patches with specific topologies. For advanced computer vision applications, e.g., object categorization, it is difficult to extract topological structures of real images with complex backgrounds. Therefore, we combine CTPO with EBIM [14]. The latter one has been demonstrated to be an effective computational model for object categorization. The framework of EBIM is shown in the upper part of Fig. 5. More details about EBIM can be found in [14]. The reasons why we combine these two models are listed below.

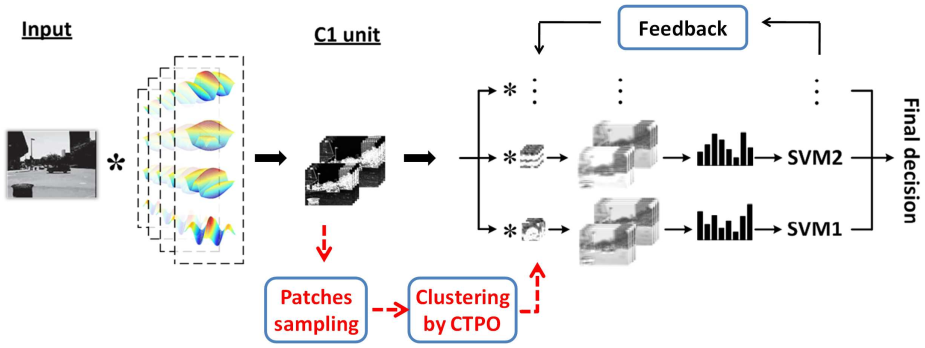


Fig. 5. The object categorization framework for experiments. The black part (the upper part) is EBIM [14] and the red part (the lower part) is CTPO. C1 unit is the processed images after being filtered by Gabor filters.

- After processed by Gabor filters in EBIM, most meaningless points in the C1 units (in Fig. 5) will be eliminated and thus we can extract topological structures efficiently and effectively.
- EBIM matches an image with a large number of patches randomly sampled from the C1 unit of training images. The random sampling technique loses its effectiveness and generalization when the number is small. Thus EBIM has to randomly draw a large number of patches with an unacceptable time cost in the training stage. Besides, meaningless patches tend to over-fit training samples. Therefore, it is necessary to keep patches with meaningful structures. Specifically, we use CTPO to extract their quotient distance histograms to cluster the original patches. Afterward, patches with meaningful topological structures are preserved and processed in the later computation. Fig. 5 shows the framework of the experiments.

In the following, we conduct object categorization experiments on two databases. The purpose of these experiments is to test the ability of CTPO in enhancing EBIM for object categorization.

MIT-CBCL Street Scene database. The MIT-CBCL street scene database [16] contains three kinds of objects: car, pedestrian and bicycle. The training and testing examples have been divided in the database for experiments.

Table 2 shows the performance comparison among our approach and C1 [17], [18], HoG [19], C1+Gestalt [18] and EBIM [14]. Our approach achieves the best performance in car and bicycle detection and is comparable to the best in pedestrian detection. The computation cost of our approach in the tests is comparable to EBIM and much less than the others. Besides, compared with EBIM, the time of training is reduced from about 20 hours to about 2 hours because our approach removes a large number of meaningless patches for matching.

Categories	Car	Pedestrian	Bicycle	time
C1	94.38	81.59	91.43	–
HoG	91.38	90.19	87.82	$\approx 0.5s$
C1+Gestalt	96.40	95.20	93.80	$> 80s$
EBIM	98.54	85.33	96.49	$\approx 0.02s$
Ours	98.80	90.28	96.86	$\approx 0.02s$

Table 1. Object detection results obtained by several state-of-the-art methods on the MIT-CBCL Street Scene database. The performance is measured in terms of EER (Equal-Error-Rate). The last column is the averaged time cost to process an image (128×128) in test.

GRAZ Database. The GRAZ database [20], built by Opelt et al., is a more complex database. To avoid the limitation that certain methods tend to emphasize background context, the backgrounds of the images in GRAZ- 02 are similar in all categories of objects. According to the depiction in [20], we strictly follow their way: 100 positive samples and 100 negative samples are randomly chosen as training samples. 50 other positive samples and 50 other negative samples are chosen as testing samples. The experimental results on the GRAZ-02 database are shown in Fig. 6 by ROC curves. Although our algorithm does not achieve the best performance compared with [21], [22], it is still comparable to them and we consider that CTPO is promising as psychophysically inspired initial research.

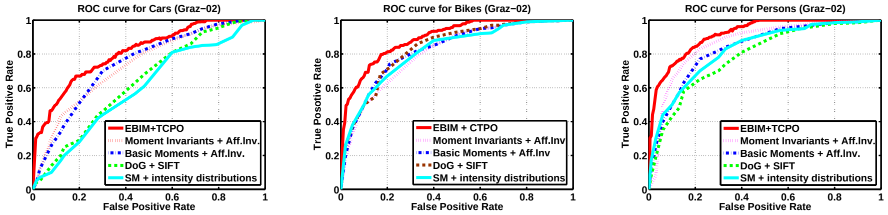


Fig. 6. (please view in color) Comparison of ROC curves on GRAZ-02 database. EBIM+TCPO is our approach. Other results are reported on [20].

5 Conclusion

In this paper, we have analyzed representative research about the primitives of visual perception. Inspired by Chen's theory of topological perceptual organization, a great breakthrough in visual psychology, we propose a computational topological perceptual organization (CTPO). The most important contribution in this paper is that we bridge the gap between Chen's psychological theory and the computer vision theory. Specifically, we have analyzed Chen's theory from the viewpoint of computational vision and building a computational model for computer vision applications. Empirical studies have proved that CTPO is consistent with Chen's theory and outperform some popular algorithms.

It is necessary to emphasis that CTPO is not designed for some specific applications but a new viewpoint to understand the organization of primal vision information in computer science. It can be easily extended to many other applications. The success of CTPO envisions the significance and potential of the early holistic registration in computer vision.

Acknowledgment

This work is supported by National Basic Research Program of China (Grant No. 2004CB318100), National Natural Science Foundation of China (Grant No. 60736018, 60723005), NLPR 2008NLPRZY-2, National Hi-Tech Research and Development Program of China (2009AA01Z318), National Science Founding (60605014, 60875021), Nanyang Technological University Nanyang SUG Grant (M58020010).

References

1. Marr, D.: Representing visual information: A computational approach Lectures on Mathematics in the Life Science, 10:61-80, 1978.
2. Marr, D.: A computational investigation into the human representation and processing of visual information San Francisco: Freeman, 1982.
3. Lowe, D.G.: Distinctive image features from dcale-invariant key-points. International Journal of Computer Vision **2**(60) (2004) 91–110
4. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence **27**(10) (2005) 1615–1630
5. Filed, D.J., Hayes, A., Hess, R.F.: Good continuation and the association field: Evidence for localfeature integration by the visual system. Vision Research **33** (1993) 173–193
6. Chen, L.: Topological structure in visual perception. Science **218** (1982) 699–700
7. Chen, L.: The topological approach to perceptual organization. Visual Cognition **12**(4) (2005) 553–638
8. Klein, F.: A comparative review of recent researches in geometry. Mathematische Annalen **43** (1872) 63–100
9. Zhuo, Y., Zhou, T.G., Rao, H.Y., Wang, J.J., Meng, M., Chen, M., Zhou, C., Chen, L.: Contributions of the visual ventral pathway to long-range apparent motion. Science **299** (2003) 417–420
10. J. Todd, e.a.: Commentaries: stability and change. Visual Cognition **12**(4) (2005) 639–690
11. Chen, L.: Reply: Author's response: Where to begin? Visual Cognition **12**(4) (2005) 691–701
12. Ling, H.B., Jacobs, D.W.: Deformation invariant image matching. ICCV (2005)
13. Ling, H.B., Jacobs, D.W.: Shape classification using the inner-distance. IEEE Trans. on Pattern Analysis and Machine Intelligence **29**(2) (2007) 286–299
14. Huang, Y.Z., Huang, K.Q., Tao, D.C., Wang, L.S., Tan, T.N., Li, X.L.: Enhanced biological inspired model. CVPR (2008)

15. Tenenbaum, J.B., Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality. Science **290**(22) (2000) 2319–2323
16. : http://cbcl.mit.edu/software-datasets/streetscenes.
17. Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., Poggio, T.: Robust object recognition with cortex-like mechanisms. IEEE Trans. on Pattern Analysis and Machine Intelligence **29**(3) (2007)
18. Bileschi, S., Wolf, L.: Image representations beyond histograms of gradients: The role of gestalt descriptors. CVPR (2007)
19. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. CVPR (2005)
20. Opelt, A., Pinz, A., Fussenegger, M., Auer, P.: Generic object recognition with boosting. IEEE Trans. on Pattern Analysis and Machine Intelligence **28**(3) (2006) 416–431
21. Leordeanu, M., Hebert, M., Sukthankar, R.: Beyond local appearance: Category recognition from pairwise interactions of simple features. CVPR (2007)
22. Ling, H., Soatto, S.: Proximity distribution kernels for geometric context in category recognition. ICCV (2007)