# SPATIAL COORDINATE CODING TO REDUCE HISTOGRAM REPRESENTATIONS, DOMINANT ANGLE AND COLOUR PYRAMID MATCH

*Piotr Koniusz and Krystian Mikolajczyk*

University of Surrey, Guildford, UK

## ABSTRACT

Spatial Pyramid Match lies at a heart of modern object category recognition systems. Once image descriptors are expressed as histograms of visual words, they are further deployed across spatial pyramid with coarse-to-fine spatial location grids. However, such representation results in extreme histogram vectors of 200K or more elements increasing computational and memory requirements. This paper investigates alternative ways of introducing spatial information during formation of histograms. Specifically, we propose to apply spatial location information at a descriptor level and refer to it as Spatial Coordinate Coding. Alternatively, x, y, radius, or angle is used to perform semi-coding. This is achieved by adding one of the spatial components at the descriptor level whilst applying Pyramid Match to another. Lastly, we demonstrate that Pyramid Match can be applied robustly to other measurements: Dominant Angle and Colour. We demonstrate state-of-the art results on two datasets with means of Soft Assignment and Sparse Coding.

***Index Terms***— Image classification, spatial pyramid match, dominant angle, colour, soft assignment, sparse coding, bags-of-words

## 1. INTRODUCTION

Spatial Pyramid Match [1] has provided foundations for majority of the modern object category recognition approaches. It was derived from Pyramid Match Kernel [2] that partitions descriptor space and immediately became a popular method to incorporate spatial information into class models. A number of systems apply this scheme, to name a few: Soft Assignment with $\chi^2$ kernel [3, 4, 5], Linear Coordinate Coding [6], Sparse Coding [7], Local Coordinate Coding [8], and approaches using Fisher Kernels [9] or Super Vector Coding [10]. Note the last two methods produce extremely large histograms which are further extended with SPM scheme to boost their performance. This results in large representation sizes up to $[(2D+1)\times K-1]\times S(N^2)$ and $(D+1)\times K\times S(N^2)$ respectively, where $D$ denotes dimensionality of applied descriptors, $K$ is a visual vocabulary size, $N$ is a number of SPM levels, and $S(N^l)=\sum_{n=1}^{N} n^l$. Another approach is based on features extracted from spatial maps derived from a hierarchical Gaussian process [11] to form Pyramid Match representation. Further, some approaches combine horizontal and vertical spatial image partitioning, e.g. [4, 12] used $1\times1, 2\times2$, and $1\times3$ horizontal, $3\times1$ vertical windows whilst [13, 10] used $1\times1, 2\times2$, and $1\times3$ horizontal divisions.

As a first contribution, we propose a scheme called Spatial Coordinate Coding that applies spatial coordinate information at the descriptor level. This reduces the histogram sizes from $H\times S(N^2)$ to $H\times S(1^1)$, where $H$ is a size of the input histograms. Further, we manipulate spatial information to be absorbed partially at the descriptor and SPM levels and reduce histogram sizes from $H\times S(N^2)$ to $H\times S(N^1)$. SCC is demonstrated to work with two popular descriptor coding methods: Soft Assignment [3, 5] and Sparse Coding [7]. Note, such scheme can be also applied with alternative coding methods [9, 10]. As a next contribution, we investigate application of Pyramid Match to different types of measurements. Dominant Angle (DA) [14] can be applied in place of spatial information by: i) DA normalising all descriptors, and ii) applying Pyramid Match to DA directly. The colour information of Segmentation-Based Descriptors [12] is experimented with in similar spirit. Lastly, we demonstrate that Spatial Coordinate Coding and Colour Pyramid Match deliver state-of-the-art results on two datasets.

## 2. SPATIAL COORDINATE CODING

Popular techniques for representing images as histograms with means of local image descriptors are Soft Assignment [3, 5] and Sparse Coding [7]. They apply spatial information at Pyramid Match level yielding long histograms of size $H\times S(N^2)$. Let $\boldsymbol{x}_n^s=[\frac{x_n^s}{w_{im}},\frac{y_n^s}{h_{im}}]^T$ be spatial coordinates of a descriptor $\boldsymbol{x}_{n\in\{1,..,N\}}$ normalised with respect to image width $w_{im}$ and height $h_{im}$. Furthermore, let $\boldsymbol{x}_n^p=[r,\phi]^T$ be vectors with the unit normalised radius $r=\sqrt{(\frac{x_n^s}{w_{im}}-\frac{1}{2})^2+(\frac{y_n^s}{h_{im}}-\frac{1}{2})^2}/\frac{\sqrt{2}}{2}$ and angle $\phi=[\phi(\frac{x_n^s}{w_{im}}-\frac{1}{2},\frac{y_n^s}{h_{im}}-\frac{1}{2})+\pi]/(2\pi)$. Let $\boldsymbol{m}_{k\in\{1,..,K\}}$ be visual words of a vocabulary of size $K$ built by either k-means or randomly sampling the descriptors of a given training set (Random Descriptor Set Sampling aka RDSS). Let $\boldsymbol{m}_k^s$ and $\boldsymbol{m}_k^p$ be the corresponding spatial vocabulary information for $(x,y)$ and $(r,\theta)$ parametrisations, respectively.

In order to prevent full Pyramid Match and to benefit from the spatial information, we propose to extend Soft Assignment and Sparse Coding schemes by applying into their workings either: i) one of spatial parametrisations $\boldsymbol{x}_n^{'}$ such as $\boldsymbol{x}_n^s/2$ or $\boldsymbol{x}_n^p/2$ leading to histogram sizes $H \times S(1^1)$, or ii) one of semi-spatial parametrisations $x_n^{'}$ such as $\frac{x_n^s}{w_{im}}$, $\frac{y_n^s}{h_{im}}$, $r$, or $\phi$. In the latter case, complementary spatial channels, e.g. $x$ and $y$ are processed one by SCC and the other by SPM scheme. The same holds if one chooses $r$ and $\theta$. This leads to smaller size $H \times S(N^1)$ compared to standard SPM: $H \times S(N^2)$.

**Spatial Coordinate Coding for Soft Assignment.** Soft Assignment [3, 5] is derived from GMM [15] with simplified model parameters $\theta = (\theta_1, ..., \theta_K) = ((\boldsymbol{m}_1, \sigma), ..., (\boldsymbol{m}_K, \sigma))$. $K$ denotes number of components, $\boldsymbol{m}_{k\in\{1,..,K\}}$ are Gaussian means, $\sigma$ is the smoothing factor, and $\boldsymbol{x}_{n\in\{1,..,N\}}$ are the descriptors of a dataset. The Component Membership Probability for this model is expressed as:

$$p(k|n) = \frac{g(\boldsymbol{x}_n; \boldsymbol{m}_k, \sigma)}{\sum_{k'=1}^{K} g(\boldsymbol{x}_n; \boldsymbol{m}_{k'}, \sigma)} \tag{1}$$

The histogram representation is expressed as the expected value of membership probabilities per component $k$ over descriptors $\boldsymbol{x}_{n\in N_{im}}$ of an image $im$: $[E_{N_{im}}(p(k|n))]_{k\in\{1,..,K\}}$.

Enhancing formula 1 with spatial or semi-spatial information can be done by adding spatially parametrised vectors $\boldsymbol{x}_n^{'}$ and $\boldsymbol{m}_k^{'}$ to Gaussian components as follows:

$$g^{'}(n,k) = g[(1-\alpha)\boldsymbol{x}_n; (1-\alpha)\boldsymbol{m}_k, \sigma^{'}]g(\alpha\boldsymbol{x}_n^{'}; \alpha\boldsymbol{m}_k^{'}, \sigma^{'}) \tag{2}$$

The additional parameter $\alpha \in\!< 0, 1)$ balances the strength of spatial coordinates versus descriptor vectors. Redefined membership probabilities are expressed by:

$$p(k|n) = \frac{g^{'}(n,k)}{\sum_{k'=1}^{K} g^{'}(n,k')} \tag{3}$$

Optimal smoothing factor $\sigma^{'}$ of Soft Assignment reformulated in equation 3 differs from $\sigma$ due to the additional spatial information introduced to the model. However, there is a relation between $\sigma$ and $\sigma^{'}$ that can be approximated as:

$$\sigma^{'} \approx \sigma\Big(1 + \frac{\sqrt{d}\alpha^2}{1-\alpha}\frac{d}{D}\Big) \tag{4}$$

We skip derivations of equation 4, though, it suffices to say $\sigma$ is extrapolated proportionally to the increase in both descriptor dimensionality and energy introduced by adding spatial information. For a pair $(x, y)$ or $(r, \phi)$ (Spatial Coordinate Coding) $d = 2$. For Semi-Spatial Coordinate Coding $d = 1$. $D$ is the descriptor dimensionality. One can either extrapolate $\sigma^{'}$ from $\sigma$ or estimate it from the data [5].

**Spatial Coordinate Coding for Sparse Coding.** The operating principle of Sparse Coding [7] is to express each descriptor vector as a sparse linear combination of neighbouring dictionary anchors. First norm over assignments favours only a small subset of activations leading to sparsity. This was found to perform well if combined with Spatial Pyramid Match and the maximum pooling [7]. Finding sparse assignments over a given descriptor $\boldsymbol{x}_n$ and a visual vocabulary $\boldsymbol{M}_{D\times K}$ is achieved by optimising the following with respect to $\boldsymbol{u}_n$:

$$\min_{\boldsymbol{u}_n} \Big\|\boldsymbol{x}_n - \boldsymbol{M}\boldsymbol{u}_n\Big\|^2 + \beta|\boldsymbol{u}_n| \tag{5}$$

$\beta$ regulates the sparsity of the solution. The histogram representation is expressed as a maximum value of assignments per anchor $k$ over descriptors $\boldsymbol{x}_{n\in N_{im}}$ of an image $im$. Enhancing formula 5 with spatial or semi-spatial information requires adding spatially parametrised vectors $\boldsymbol{x}_n^{'}$ and $\boldsymbol{m}_k^{'}$ to Lasso problem:

$$\min_{\boldsymbol{u}_n}(1-\alpha)\Big\|\boldsymbol{x}_n - \boldsymbol{M}\boldsymbol{u}_n\Big\|^2 + \alpha\Big\|\boldsymbol{x}_n^{'} - \boldsymbol{M}^{'}\boldsymbol{u}_n\Big\|^2 + \beta|\boldsymbol{u}_n| \tag{6}$$

Note, both Soft Assignment (equation 1) and Sparse Coding (equation 5) can be enhanced by Spatial Coordinate Coding by just concatenating appropriately image descriptors with the corresponding spatial representation $\boldsymbol{x}_n^{'}$, i.e.: $\boldsymbol{x}_n^a = [\sqrt{1-\alpha}\boldsymbol{x}_n^T, \sqrt{\alpha}(\boldsymbol{x}_n^{'})^T]^T$. The same applies to $\boldsymbol{m}_k$.

## 3. DOMINANT ANGLE AND COLOUR PYRAMID MATCH

This section provides details on how to exploit Dominant Angle [14] (DA) and colour [12] information in Pyramid Match. Variety of cues contribute valuable information and may be appropriate for quantising them at multiple levels. The sun and clouds appear in the sky, thus they are mostly contained in the upper parts of images. If spatial positions $X_s$ of an object $s$ introduce a spatial bias in images such that $p(o = s|\boldsymbol{x}) \geq p(o = s)$ for $\boldsymbol{x} \in X_s$, then the orientation of dominant edges within images should also induce an orientation bias. DA is a direction with respect to the origin of a local image descriptor indicated by the highest gradients within the descriptor. Note, trunks of trees $t$ and fences are more likely to maintain vertical positions $\Theta_t$, therefore $p(o = t|\theta) \geq p(o = t)$ if $\theta \in \Theta_t$. Facial complexion $f$ or fur of animals are likely to be of a limited colour set $C_f$, thus $p(o = f|\boldsymbol{c}) \geq p(o = f)$ if $c \in C_f$ is observed. Interestingly, using rotationally variant descriptors with bag-of-words yields much better results [12] compared to rotationally invariant counterparts. This shows that the rotational bias helps in recognition. We introduce DA to the classification process in two ways: i) by setting $x_n^{'} = \theta_n$, or ii) by performing Pyramid Match directly on $\theta_n$. Regarding colour, Segmentation-Based Descriptors [12] were used as they consist of: i) orientations of image gradients $\boldsymbol{x}^o$, ii) eigenvalues $\boldsymbol{x}^e$ of dominant shapes, iii) opponent colour histograms $\boldsymbol{x}^c$. We reduced 20D opponent vectors by PCA to 10D. Colour component $c$ with the highest variance was fed to Pyramid Match. The remaining 9 components replaced the original opponent vectors (20D reduced to 9D).

| $SC_{1234}$ | $SC$+$SCC$ | $SA_{123}$ | $SA$+$SCC$ | $SA$+$SCC$ |
|---|---|---|---|---|
| Lin+SVM | Lin+SVM | $\chi^2$+KDA | $\chi^2$+KDA | $\chi^2$+KDA |
| 1ker+val | 1ker+val | 1ker+val | 1ker+val | multiker+tst |
| 48.7 | 47.0 | 49.8 | **51.6** | **62.15** |

**Table 1**. MAP for Pascal 2010 Action Classification.

| DA Inv. | DA Var. | DACC$\alpha$=$\frac{1}{2}$ | DACC$\alpha$=$\frac{2}{3}$ | DACC$\alpha$=$\frac{4}{5}$ |
|---|---|---|---|---|
| 46.00 | 50.23 | 47.2 | 49.80 | **50.24** |
| DA Var.+ | $DA_{12468}$ | $DA_{136912}$ | $DA_{136912}$ | |
| SPM 54.3 | 52.30 | **53.40** | SPM **56.3** | |

**Table 2**. MAP for Pascal 2007 Main Challenge comparing impact of Dominant Angle on classification.

## 4. EVALUATIONS AND CLASSIFICATION RESULTS

This section provides an experimental insight regarding Spatial Coordinate Coding versus Spatial Pyramid Match [4, 12]. Tests were performed on Pascal 2010 [16] Action Classification set (301 training, 307 validation, and 613 testing bounding boxes) and Flower 17 [17] set (3 splits of data, each consisting of 680 training, 340 validation, and 340 testing images). For Pascal 2010, we report results mainly on validation set as testing set is not publicly available. We also quote our results on test set submitted for Pascal 2010 [16] competition. Experiments on Dominant Angle Pyramid Match were performed on Pascal 2007 [16] Main Challenge.

Two variants of descriptors were exploited: grey-scale SIFT [14] (Pascal 2010 and 2007 sets) and Segmentation-Based Image Descriptors [12] (Flower 17 set). Dense feature sampling on a regular grid with the intervals of 8, 14, 20, and 26 pixels, and patch radii of 16, 24, 32, and 40 pixels was applied for SIFT. This produced 1200, 3690, and 2300 vectors per image on average on Pascal sets and Flower 17 set respectively. We observed KDA [4] classifier always worked better with $\chi^2$ [4] and SVM with linear kernels, thus we used such set-up. Kernels were formed from either soft assigned [3, 5] (SA) or sparsely coded [7] (SC) histograms. As a reference, Spatial Pyramid Match (SPM) with 3 and 4 levels of depth was employed for SA and SC respectively. The visual vocabulary of size $K = 4000$ was produced by k-means (Pascal 2010 and 2007) and random sampling of descriptors on the training set (Random Descriptor Set Sampling aka RDSS) of Flower 17. Note, we are not concerned directly with optimising visual dictionaries as our ideas are not affected by them.

**Pascal 2010 and Spatial Coordinate Coding.** Pascal 2010 Action Classification provides bounding boxes delineating humans performing actions to classify. Every person's head is roughly aligned to the top middle location of such bounding box. Positions of interacted objects can be expressed with respect to the top middle reference point. Thus, Spatial Coordinate Coding is applied and compared with Spatial Pyramid Match (SPM). Table 1 presents the results achieved on this set. Sparse Coding $SC_{1234}$ with SPM (4 levels) turned out worse than Soft Assignment $SA_{123}$ with SPM (3 levels). Also, Spatial Coordinate Coding combined with Sparse Coding $SC$+$SCC$ seemed slightly worse than $SC_{1234}$. SA with SCC (denoted as $SA$+$SCC$) was the strongest performer leading to state-of-the-art **62.15**% MAP on testing set compared to other systems [16]. This was achieved by averaging multiple kernels of descriptor variants as in [12, 4]. We observed SA with $\chi^2$ is well suited to benefit from SCC scheme.

**Pascal 2007 and Dominant Angle Pyramid Match.** Pascal 2007 consists of 20 object categories with high variability in intra-class appearance, rotation, and spatial position. This section presents impact of Dominant Angle (DA) combined with Pyramid Match on classification. The following results were achieved with Soft Assignment, $\chi^2$ kernel, and KDA classifier. According to table 2, DA is an important modality for robust classification. DA Inv. is a baseline result with SIFT descriptors deemed invariant to rotation. Applying invariance decreased performance compared to Dominant Angle variant set (DA Var.) from 50.23% MAP down to 46% MAP. We further used DA invariant SIFT and injected DA directly to equations 2 and 3 (referred to as DACC) with $\alpha = \frac{1}{2}$, $\frac{2}{3}$, and $\frac{4}{5}$. DACC$\alpha$=$\frac{4}{5}$ achieved results of 50.24% MAP on a par with DA Var. Therefore, DA is a robustly estimated reliable cue. After reintroducing it back to the pipeline, full performance was regained. Further, DA is also suited for quantisation at multiple levels with Pyramid Match. Note, the angle invariant SIFT fed to Pyramid Match with 5 levels of DA splits 1, 3, 6, 9, 12 ($DA_{136912}$ in the table) achieved 53.4% MAP that outperformed DA Var. SIFT by **3.1**% due to multiple levels of angle quanitsation. This combined with Spatial Pyramid Match (DA Var.+SPM) boosted performance from 54.3% to **56.3**% MAP with one set of descriptors.

**Flower 17, Spatial Coordinate Coding, and Colour Pyramid Match.** Performance of both Spatial and Semi-Spatial Coordinate Coding was evaluated in more detail on Flower 17 dataset with means of both Soft Assignment and Sparse Coding. According to results in table 3, Soft Assignment with SVM and the linear kernel (SA Lin SVM row) achieved better results of 84.7% MAP if using full Spatial Pyramid Match (SPM) with 3 levels of depth rather than SCC. Radius and $\theta$ parametrised SPM (SPM$r\theta$) was a close performer. Also, Spatial Coordinate Coding SCC$\alpha$=$\frac{9}{14}$ achieved close results of 82.93% MAP. The gap of 1.8% in performance between two methods is bridged by Semi-Spatial Coordinate Coding (table 4). Note, SA with $\chi^2$ and KDA classifier (SA $\chi^2$ KDA row) exploited SCC to its fullest potential outperforming SPM ($N = 3$) by roughly **1.8**% and reducing histogram sizes from $4K \times S(3^2) = 56K$ to $4K \times S(1^1) = 4K$. Lastly, Sparse Coding with the linear kernel and SVM (SC Lin SVM row) benefited **1.2**% from SCC over no spatial information added (NO SCC) whlist SPM ($N = 4$) led to about 1.6% over NO SCC. These results are further improved by Semi-Spatial Coordinate Coding as presented in table 4.

According to table 4 (first row), all semi-spatial combinations improved results by up to 0.7% over SA with the

| | | | | |
|---|---|---|---|---|
| SA Lin SVM | SCC$\alpha$=$\frac{6}{11}$ 82.36 | SCC$\alpha$=$\frac{9}{14}$ 82.93 | SPM 84.7 | SPM$r\theta$ 83.7 |
| SA $\chi^2$ KDA | SCC$\alpha$=$\frac{6}{11}$ 90.96 | SCC$\alpha$=$\frac{9}{14}$ **91.16** | SPM 89.3 | SPM$r\theta$ 89.63 |
| SC Lin SVM | NO SCC$\alpha$=0 87.16 | SCC$\alpha$=$\frac{1}{3}$ **88.43** | SPM 88.86 | |

**Table 3**. MAP for Flower 17 set comparing Spatial Pyramid Match with Spatial Coordinate Coding scheme.

linear kernel and SPM. We combine SPM and SCC in the table with specific semi-spatial channels $x$, $y$, $r$, $\theta$ as in section 2. SA with $\chi^2$ kernel and KDA classifier (second row) favours full SCC reaching **91.16**% compared to 90.4% MAP for SPM$y$+SCC$x$. Sparse Coding (bottom row) benefited from semi-spatial variants SPM$x$+SCC$y$ and SPM$\theta$+SCC$r$ outperforming full SPM by 0.34% and limiting histogram sizes from $4K \times S(4^2) = 120K$ to $4K \times S(3^1) = 24K$.

As experiments on Flower17 benefit from colour cues on the descriptor level [12], we also investigated benefits of Pyramid Match quantisation applied to the colour as explained in section 3. Soft Assignment with $\chi^2$ kernel, KDA classifier, and SCC (**91.16**% MAP, **86.4**% accuracy) were further enhanced by this method and yielded state-of-the-art **92.2**% MAP (**87.4**% accuracy). This, if combined with Colour SIFT at kernel level [4], increased to **95.2**% MAP (**91.4**% accuracy). In contrast, the runner-up reports 86.7% accuracy [18].

## 5. CONCLUSIONS

We have presented a novel method injecting spatial information to the classification process at the descriptor level. This resulted in significantly smaller histogram representations and improved performance for Soft Assignment and $\chi^2$ kernels. Also, semi-spatial approach was proposed to benefit more demanding histogram coding approaches like Sparse Coding with linear kernels. Overlooked importance of Dominant Angle mechanism was brought to attention as we demonstrated its benefit on classification if applied to Pyramid Match and complemented by Spatial Pyramid Match. As objects exhibit variable intra-class colour similarity, we showed that colour components also thrive on quantising with Pyramid Match. This led us to state-of-the-art results on both Pascal 2010 Action Classification and Flower17 datasets.

**Acknowledgements.** This work was sponsored by the BBC Future Media and Technology and EPSRC EP/F003420/1 research grants.

## 6. REFERENCES

[1] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," *CVPR*, vol. 2, pp. 2169–2178, 2006.

[2] K. Grauman and T. Darrell, "The pyramid match kernel: Discriminative classification with sets of image features," *ICCV*, pp. 1458–1465, 2005.

| | | | | |
|---|---|---|---|---|
| SA Lin SVM | SPM$y$+ SCC$x$ **85.4** | SPM$x$+ SCC$y$ **85.2** | SPM$\theta$+ SCC$r$ **85.5** | SPM$r$+ SCC$\theta$ **85.5** |
| SA $\chi^2$ KDA | SPM$y$+ SCC$x$ 90.4 | SPM$x$+ SCC$y$ 90.1 | SPM$\theta$+ SCC$r$ 90.2 | SPM$r$+ SCC$\theta$ 90.2 |
| SC Lin SVM | SPM$y$+ SCC$x$ 88.8 | SPM$x$+ SCC$y$ **89.2** | SPM$\theta$+ SCC$r$ **89.0** | SPM$r$+ SCC$\theta$ 88.6 |

**Table 4**. MAP for Flower 17 set utilising Semi-Spatial Pyramid Match.

[3] J. C. van Gemert, C. J. Veenman, A. W. M. Smeulders, and J. M. Geusebroek, "Visual word ambiguity," *PAMI*, 2010.

[4] M.A. Tahir, J. Kittler, K. Mikolajczyk, F. Yan, K. van de Sande, and T. Gevers, "Visual category recognition using spectral regression and kernel discriminant analysis," *ICCV Workshop*, 2009.

[5] P. Koniusz and K. Mikolajczyk, "Soft assignment of visual words as linear coordinate coding and optimisation of its reconstruction error," *ICIP*, 2011.

[6] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," *NIPS*, 2009.

[7] J. Yang, K. Yu, Y. Gong, and T. S. Huang, "Linear spatial pyramid matching using sparse coding for image classification.," *CVPR*, pp. 1794–1801, 2009.

[8] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, "Locality-constrained linear coding for image classification," *CVPR*, 2010.

[9] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," *ECCV*, pp. 143–156, 2010.

[10] X. Zhou, K. Yu, T. Zhang, and T. S. Huang, "Image classification using super-vector coding of local image descriptors," *ECCV*, pp. 141–154, 2010.

[11] X. Zhou, N. Cui, Z. Li, F. Liang, and T. S. Huang, "Hierarchical Gaussianization for Image Classification," *ICCV*, 2009.

[12] P. Koniusz and K. Mikolajczyk, "On a quest for image descriptors based on unsupervised segmentation maps," *ICPR*, pp. 762–765, 2010.

[13] M. Marszalek, C. Schmid, H. Harzallah, and J. Van De Weijer, "Learning object representations for visual object class recognition," *ICCV Workshop*, 2007.

[14] D. G. Lowe, "Object recognition from local scale-invariant features," *CVPR*, vol. 2, pp. 1150–1157, 1999.

[15] J. Bilmes, "A gentle tutorial of the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models," Tech. Rep., 1998.

[16] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results," http://pascallin.ecs.soton.ac.uk/challenges/VOC, 2010.

[17] M. E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," *ICCV*, pp. 722–729, 2008.

[18] F. Yan, K. Mikolajczyk, M. Barnard, H. Cai, and J. Kittler, "Lp norm multiple kernel fisher discriminant analysis for object and image categorisation," *CVPR*, 2010.