Open access • Proceedings Article • DOI:10.1109/CVPR.2011.5995637

Multi-spectral SIFT for scene category recognition — Source link

Matthew Brown, Sabine Süsstrunk

Institutions: École Polytechnique Fédérale de Lausanne

Published on: 20 Jun 2011 - Computer Vision and Pattern Recognition

Topics: Contextual image classification and Multispectral image

Multi-spectral SIFT for Scene Category Recognition

Matthew Brown and Sabine Süsstrunk
School of Computing and Communication Sciences,
École Polytechnique Fédérale de Lausanne (EPFL).
{matthew.brown,sabine.susstrunk}@epfl.ch

Abstract

We use a simple modification to a conventional SLR camera to capture images of several hundred scenes in colour (RGB) and near-infrared (NIR). We show that the addition of near-infrared information leads to significantly improved performance in a scene-recognition task, and that the improvements are greater still when an appropriate 4-dimensional colour representation is used. In particular we propose MSIFT – a multispectral SIFT descriptor that, when combined with a kernel based classifier, exceeds the performance of state-of-the-art scene recognition techniques (e.g., GIST) and their multispectral extensions. We extensively test our algorithms using a new dataset of several hundred RGB-NIR scene images, as well as benchmarking against Torralba's scene categorization dataset.

Figure 1: Examples from our database of RGB-NIR images. Notice that the NIR band exhibits noticeable differences at the scene level: sky and water are dark, foliage is bright, and details are more clearly resolvable in haze.

1. Introduction

Silicon based digital cameras are naturally sensitive to near-infrared (NIR) light, but are prevented from capturing it by a filter known as a "hot-mirror" between the lens and the CCD. It has been argued that removing this limitation and devoting a fraction of the pixels to NIR [12] could be beneficial in computational photography applications (e.g., dehazing [22] and Dark Flash Photography [10]). Recent applications have also demonstrated the utility of near-infrared in image understanding, for example, Microsoft's Kinect system, which uses active NIR illumination to estimate scene depths. In this work, we argue that *passive* NIR can also be useful in Computer Vision. To demonstrate this, we choose the application of scene recognition, and will aim to exploit the material differences between scene elements in NIR and RGB [21] to improve recognition performance (see Figure 1).

Scene recognition is a long-standing problem in computer vision, being an important element in contextual vision [25, 9]. Scene recognition capabilities are also starting to appear in digital cameras[2], where "Intelligent Scene Recognition" modules can help to select appropriate aperture, shutter speed, and white balance.

A benchmark approach to computational scene recognition was developed by Oliva and Torralba [18]. Their GIST descriptors, a succinct summary of spatial frequencies and their arrangement, was inspired by the rapid categorisation and coarse to fine processing believed to feature in human vision [23]. Riesenhuber and Poggio's HMAX models similarly attempt to mimic the processing in V1, and variations of this so called "Standard Model" have also been successful in category recognition problems [16, 19]. Local feature methods are also very popular in category recognition [13], recent work has extended these methods to effectively make use of colour [26].

Scene recognition has been of particular interest to visual psychologists and neuroscientists. One intriguing aspect is that it can be accomplished extremely rapidly in human vision [23]. This fact has led to much debate and investiga-

[2]Sony W170, Nikon D3/300

tion into the visual processes that might occur. For example, Fei-Fei et al. [4] have argued that the absence of colour fails to make scene recognition more attention demanding, whereas Oliva and Schyns and Goffaux et al. [17, 8] find that reaction times are improved if diagnostic scene colours are preserved.

In computer vision applications, effective use of colour for recognition may require illumination estimation [6, 5] or computing invariants [7] under complex illumination conditions. Correlations between the colour bands are strong, and since the luminance (greyscale) component amounts to around 90% of the signal energy, many practitioners ignore colour entirely. One attraction of looking at near-infrared is that it has a much weaker dependence on R, G and B than they do to each other, which should amplify any gains from effective multispectral techniques.

In related literature, researchers have studied the statistics of images in the *far*-infrared (4-12μm) [14], as well as demonstrating enhanced pedestrian tracking using this band [27]. These applications require a specialist detector that is dedicated for use in the far-infrared band. In this work we focus instead on the *near*-infrared (750-1100nm), that can be captured using an ordinary digital camera. In principle (by using a camera with a modified Bayer pattern), NIR pixels could be captured jointly with RGB [12].

2. Contribution

The main contributions of our work are:

1. MSIFT: a multispectral SIFT descriptor that effectively uses the information in multiple spectral bands.

2. A new dataset of 477 registered colour (RGB) and near-infrared (NIR) image pairs[1].

We also conduct further investigations into existing colour SIFT descriptors, and suggest practical improvements.

3. RGB-NIR Imaging

The CCD and CMOS chips present in digital cameras are sensitive over a range of approximately 350-1100nm. Whereas human sensitivity drops off sharply at around 700nm, silicon is actually more sensitive in this region. For this reason a specific infrared blocking filter is used in addition to the red, green and blue colour filter array (CFA) elements, to prevent an unwanted NIR response. If this filter is removed, the RGB CFA elements give easily measurable responses in the near-infrared range (see Figure 2).

In this work, we use several digital SLR cameras that have been modified to remove the infrared blocking filter.

[1] See http://ivrg.epfl.ch/supplementary_material/cvpr11/index.html

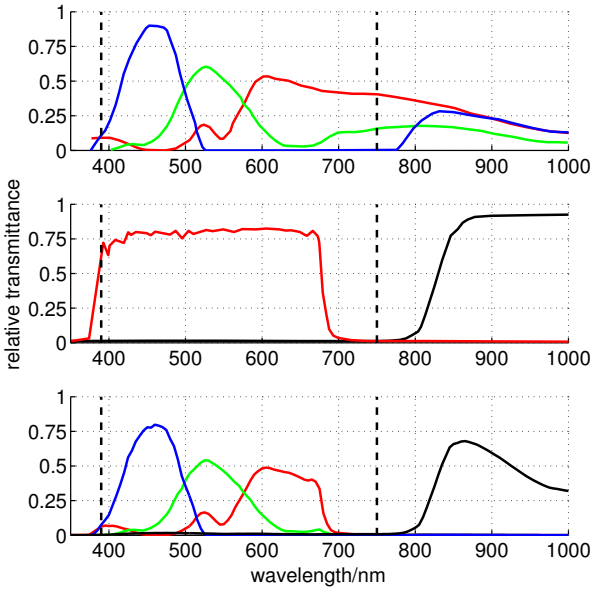


Figure 2: Silicon sensitivity extends beyond the visible spectrum (top row). We use visible and NIR pass filters (middle row) to simulate a 4-band sensor response using a conventional camera (bottom row).

Though setups involving beam splitters exist, no portable RGB-NIR still camera is yet available. Hence we capture two separate exposures, using RGB and NIR filters passing wavelengths below and above 750nm respectively. The RGB sensor responses for the visible capture are

$$\rho_i = \int_\lambda E(\lambda)S(\lambda)R_i(\lambda)F^{VIS}(\lambda)d\lambda \qquad (1)$$

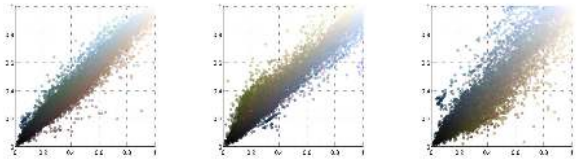
where $E(\lambda)$ and $S(\lambda)$ are the illuminant spectral power and surface spectral reflectance, $R_i(\lambda),\ i \in \{R, G, B\}$ are the sensor quantum efficiencies, and $F^{VIS}(\lambda)$ is the visible pass filter response. We form a single NIR channel by summing the responses of the same CFA elements modulated by the NIR filter

$$\rho_{NIR} = \sum_i \int_\lambda E(\lambda)S(\lambda)R_i(\lambda)F^{NIR}(\lambda)d\lambda \qquad (2)$$

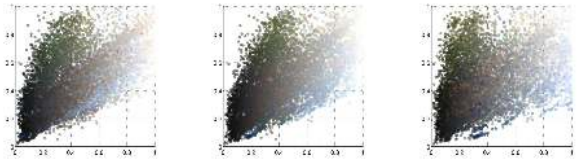
where $F^{NIR}(\lambda)$ is the spectrum of the NIR pass filter (see Figure 2).

3.1. Image Registration

Small movements of the tripod may result in a small offset between the RGB and NIR image captures. To correct for this, we use a feature based alignment algorithm [24] with robust estimation of a similarity motion model. We reject image matches with less than 50 consistent feature



(a) R-G (4.13) (b) B-G (4.26) (c) R-B (4.60)



(d) R-NIR (5.29) (e) G-NIR (5.30) (f) B-NIR (5.44)

Figure 3: Pairwise distributions of R, G, B and NIR pixels sampled from 10,000 examples in 100 images. The joint entropy of each distribution (64 bins, max entropy of 6 bits) is shown in parenthesis.

matches (this occurred in only 6 instances). Our subjects are primarily static scenes, although occasionally image motion causes small differences between the RGB and NIR bands (around 50 of 477 cases), to which our recognition algorithms will need to be robust.

4. RGB-NIR Statistical Dependencies

Figure 3 shows scatter plots of 10,000 sampled RGB-NIR pixels from our dataset, plotted as pairwise distributions in order of increasing entropy. Joint entropy is computed as $H = \sum_{ij} -p_{ij} \log_2(p_{ij})$, using a discretisation of 8 bins per dimension. Note that the visible colour entropies (R-G, R-B and B-G) are all less than the joint entropy of visible and NIR (R-NIR, G-NIR, B-NIR), with the largest entropy occurring for the spectral extremes (R-B and B-NIR). These plots suggest that NIR gives significantly different information from R, G and B, in the following we investigate whether this can be effectively used in a recognition context.

5. Multispectral SIFT

In their paper on colour SIFT descriptors [26], Van der Sande et al. noted that using opponent colour spaces gave significantly better performance in object and scene recognition than computing descriptors in the R, G and B colour channels directly. Opponent colours are present in human vision, where early processing splits the input into achromatic (luminance) and opponent (red-green, blue-yellow) parts. This can be explained in terms of efficient coding – opponent colours decorrelate the signal arriving at the L, M and S photoreceptors [2, 20]. To extend the opponent colour idea to RGB-NIR, we make use of the same idea, decorrelating the 4-dimensional RGB-NIR colour vector $\mathbf{c} = [r, g, b, i]$ by computing the eigenvectors of the covariance matrix

$$\mathbf{\Sigma_c} = \sum_k (\mathbf{c}_k - \mathbf{m_c})(\mathbf{c}_k - \mathbf{m_c})^T = \mathbf{W \Lambda W}^T \quad (3)$$

To ensure that the output remains as the input within the 4-d unit cube we apply a further linear mapping

$$c'_i = \frac{1}{\sum_j |w_{ij}|} \sum_j w_{ij} c_j - \frac{\sum_j w_{ij}^{(-)}}{\sum_j |w_{ij}|} \; . \quad (4)$$

This facilitates downstream processing, which expects intensity values in the 0-1 range. Each component of the resulting colour vector $\mathbf{c}' = [c'_1, c'_2, c'_3, c'_4]$ is thus a linear (scale and offset) transform of the decorrelated components. The raw RGB-NIR PCA components are shown in Figure 4. Similar to 3-dimensional colour, the first component is almost achromatic, containing approximately equal amounts of R, G, B, and a slightly smaller NIR contribution. The second component takes the difference of the spectral extremes (NIR and blue), which from Figure 3 are the most statistically independent. Subsequent components involve further spectral differences, and account for small fractions of the overall signal energy (see Figure 6).

To form a multispectral SIFT keypoint, we first detect difference of Gaussian extrema in the luminance component, and then form 4×4 histograms of gradient orientations $\mathbf{d}_i, \; i \in \{1..4\}$ for each channel independently. Note that since the decorrelated bands (other than the first) consist of colour differences, the gradients consist of both spatial and chromatic differences (see Figure 5). We normalise each colour band independently, which equalises the weighting of the colour gradient signals, and concatenate to form the final descriptors. Since this can be high dimensional (512 dimensions for RGB-NIR descriptors), we reduce dimensions using PCA, leading to a length n_d descriptor

$$\mathbf{d} = \mathbf{U}^T [\tilde{\mathbf{d}}_1, \tilde{\mathbf{d}}_2, ..., \tilde{\mathbf{d}}_n] \quad (5)$$

where $\tilde{\mathbf{d}}_i$ is the "clip-normalised" [11] SIFT descriptor in the ith band, and $\mathbf{U}$ is a $n_d \times 128n$ orthogonal matrix (typically $n_d \approx 128$).

6. Standard Models

As a baseline for comparison, we also test multispectral versions of two standard models for category/scene recognition:

GIST We compute Gabor filters at 3 scales and 8 orientations per scale. The image is first pre-filtered to nor-

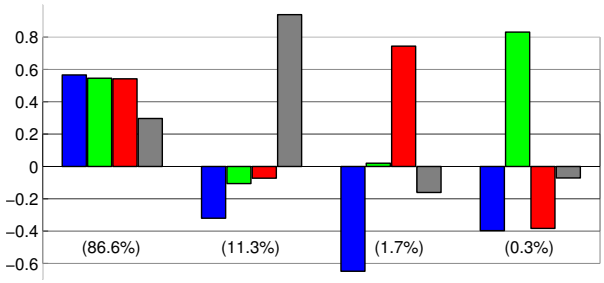


Figure 4: RGB-NIR "opponent" colours. The coloured bars show the amounts of red, green, blue and near-infrared in each PCA component. Note that the first two components (approximately achromatic and difference of B-NIR) make up 98% of the signal energy.

Figure 5: Colour parts of the derivative filters used for MSIFT (note that there is an additional near-infrared component that is not visualised)

Figure 6: RGB-NIR opponent colour components of a scene. The colour transformation converts colour differences into spatial patterns, which are well characterised by local descriptors such as SIFT. Note that there is visibly less energy in the later components.

malise the local contrast, and the descriptors are concatenated for each band. We use the implementation from Torralba [18].

HMAX Inspired by the "standard model" for recognition in visual cortex, this consists of a hierarchy of filtering and max-pooling operations. HMAX descriptors are computed independently and concatenated for each band. We use the CNS implementation by Mutch et al. [15].

The HMAX filters were trained offline using a database of 10,000 images. GIST descriptors were normalised per band, which also gave improved performance versus global normalisation. In addition to our multispectral SIFT descriptors, we also compare against the best performing colour descriptor designs proposed by Van de Sande et al. [26, 3], including opponent-SIFT, C-SIFT and rg-SIFT.

7. Database

Our test database consists of 477 images distributed in 9 categories as follows: Country(52), Field(51), Forest(53), Mountain(55), Old Building(51), Street(50), Urban(58), Water(51). The images were processed using automatic white balancing for the RGB components, and equal weights on the RGB sensor responses for the NIR components, with standard gain control and gamma correction applied.

8. Classification Experiments

To perform classification based on MSIFT descriptors, we adopt the method of Boiman et al. [1]. This assumes that the descriptors $\mathbf{z}_i$ in image I are i.i.d. Gaussian, with the class conditional density being approximated by the nearest-neighbour

$$p(I|c) = \prod_i p(\mathbf{z}_i|c) \approx \prod_i \mathrm{N}(\mathbf{z}_i, \mathrm{NN}_c(\mathbf{z}_i), \sigma^2) \quad (6)$$

where $\mathrm{NN}_c(\mathbf{z}_i)$ is the nearest neighbour of descriptor z_i in class c of the training set. We then use a Bayes classifier, choosing the class $c^* = \arg\max_c p(c|I)$ with equal class priors. For the GIST and HMAX descriptors we use Linear SVMs (C-SVC) with a constant c parameter of 100.

9. Experiments (RGB-NIR dataset)

We perform scene recognition on our dataset of 477 images, randomly selecting 99 images for testing (11 per category) and training using the rest. In all our experiments we repeat using 10 trials with a randomly selected training/test split, and quote the mean and standard deviation of the recognition rate (fraction of correct matches).

Firstly, we experiment with various colour representations for each lifting algorithm. These are: **l** = luminance (greyscale), **li** = luminance + NIR, **rgb** = red, green and blue, **rgbi** = RGB + NIR, **opp** = opponent colour space (as used for opponent-SIFT), **nopp** = normalised opponent colour space (as used for C-SIFT), **lrg** = luminance + normalised r, g (as used for rg-SIFT), **pca1** = 1^{st} RGB-NIR PCA component, **pca2** = 1^{st} and 2^{nd} RGB-NIR PCA components, **pca3** = RGB-NIR PCA components 1-3, **pca4** = RGB-NIR PCA components 1-4, **rnd** = random 4×4 linear transform (10 randomised transforms used over 10 trials each). The results are shown in Table 1. We name the combination **pca4_sift** = MSIFT (the algorithm described in section 5).

In each case there is general trend that adding more information leads to better recognition performance for all

Colour	Descriptor Algorithm		
	HMAX	GIST	SIFT
l	50.3 (±3.2)	59.9 (±3.5)	59.8 (±3.8)
li	55.9 (±3.7)	60.4 (±3.4)	64.1 (±3.6)
rgb	53.4 (±3.9)	60.0 (±3.3)	62.9 (±3.1)
rgbi	**57.1** (±4.0)	60.0 (±4.4)	**67.5** (±2.3)
opp	51.6 (±3.5)	64.2 (±3.1)	67.0 (±3.0)
nopp	52.2 (±3.3)	**64.6** (±3.6)	66.0 (±2.4)
lrg	56.4 (±3.9)	**66.3** (±3.9)	65.9 (±3.7)
pca1	51.1 (±4.3)	58.8 (±3.7)	60.6 (±2.0)
pca2	55.1 (±4.5)	62.9 (±4.3)	64.9 (±3.8)
pca3	**57.2** (±3.8)	63.1 (±3.6)	**70.0** (±2.2)
pca4	**59.2** (±2.7)	**65.9** (±2.9)	**73.1** (±3.3)
rnd	54.1 (±4.4)	58.2 (±5.2)	63.6 (±3.7)

Table 1: Recognition rates (%) for each descriptor with varying colour transforms applied. The best 3 results in each column are printed in boldface. Standard deviations over 10 test runs are in parentheses.

algorithms. The SIFT based descriptors show the greatest improvements as more information is added (59.8% for ordinary SIFT descriptors to 73.1% using MSIFT). The improvements for adding pure NIR to the greyscale band (**li**) are greater than adding unmodified RGB colour (**rgb**) for all algorithms, and in each case except GIST, the best results involve using NIR. For the GIST descriptors the best results are achieved using the **lrg** colour transform, although the results using **pca4** and both opponent (**opp** and **nopp**) colour transforms are within a single standard deviation.

Rank ordered results are shown in Figure 7. Overall, MSIFT (=**pca4_sift**) descriptors gave the best performance (73.1%), with significantly greater performance than methods that did not make use of near-infrared. The closest performing non-NIR method was **opp_sift** (67.0%). A paired t-test between these two methods gave a p-value of 0.003 (meaning 0.3% probability of observing these results if the means did not in fact differ). Generally adding any form of colour improved the results (the baseline greyscale algorithms are plotted with grey bars). As a sanity check, we also tried random 4×4 colour transforms, which performed low to mid-range within each descriptor category.

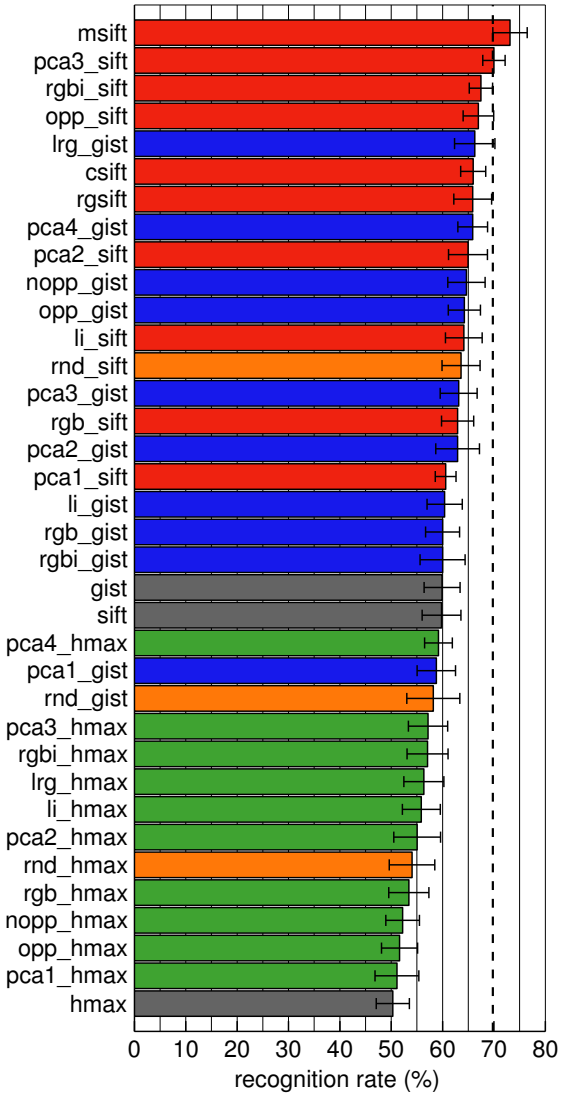


Figure 7: Rank ordered recognition rates for all algorithms. SIFT based descriptors are coloured red, GIST blue and HMAX green. The baseline greyscale algorithms are grey, and random colour transforms orange. The dotted line is the lower bound of the MSIFT confidence interval.

9.1. Dimensionality Reduction

A disadvantage with colour and multispectral SIFT descriptors is their high dimensionality, which limits scalability and increases computation time for subsequent recognition stages. We tested reducing the dimensionality of our descriptors using PCA, the results are shown in Figure 8. We found that in all cases, performance levels off at around 128 dimensions, but with the recognition rate of MSIFT around 10% higher than greyscale SIFT at this operating point. The right-hand figure demonstrates the increase in performance as colour and NIR information are added, with descriptors using colour + near-infrared (MSIFT) giving significantly better performance than those using colour only (**opp_sift**). Also, for any given dimensionality, the best performing results are those using **pca3_sift** or MSIFT (see Table 2).

We also experimented with other methods to reduce dimensions, including taking the mean, max and median of the transformed gradients instead of concatenating the spectral bands. The results were: mean 53.6% (±3.2), median

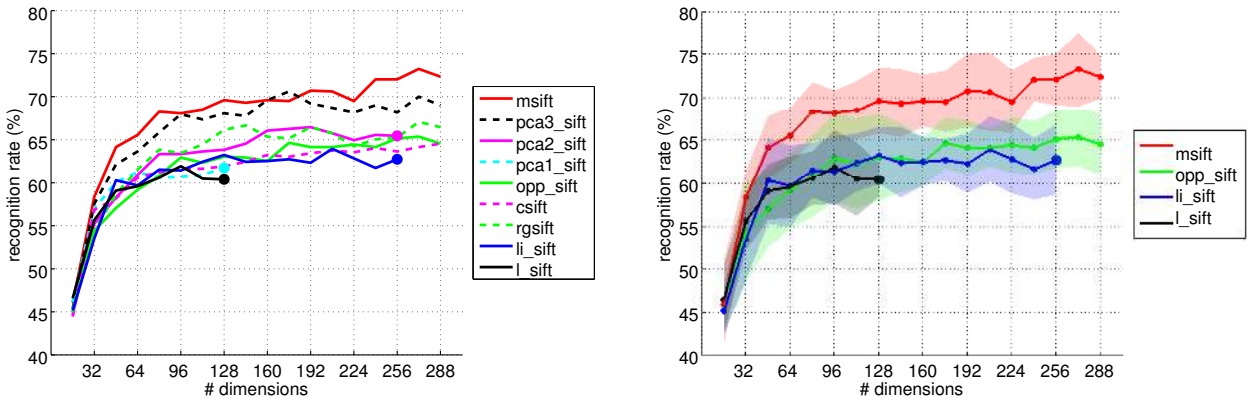


Figure 8: Performance VS number of dimensions for colour and near-infrared SIFT-based descriptors. Though the full multispectral descriptors can be high dimensional, performance plateaus, with around 128 dimensions being enough for good performance. The right-hand figure compares the performance of the best descriptors using colour plus NIR (MSIFT) and colour only (**opp_sift**), with greyscale+NIR (**li_sift**) and ordinary greyscale SIFT (**l_sift**). Note that the addition of NIR information gives significant improvements over using colour only.

Colour	Dimensions		
	64	128	256
l	59.6 (±3.6)	60.4 (±2.0)	-
li	59.7 (±4.7)	63.2 (±3.3)	62.7 (±4.1)
rgb	60.5 (±2.3)	61.2 (±2.9)	63.0 (±3.5)
rgbi	**62.6** (±4.9)	63.8 (±4.3)	64.6 (±4.0)
opp	59.2 (±4.3)	63.0 (±5.0)	65.2 (±3.3)
nopp	60.6 (±3.5)	61.9 (±3.7)	63.6 (±3.5)
lrg	61.5 (±3.9)	**66.2** (±3.4)	65.3 (±3.7)
pca1	61.2 (±3.9)	61.7 (±2.2)	-
pca2	60.9 (±2.7)	63.8 (±2.9)	**65.5** (±4.1)
pca3	**63.6** (±4.6)	**68.1** (±2.4)	**68.2** (±2.8)
msift	**65.6** (±3.2)	**69.6** (±3.8)	**72.0** (±2.9)

Table 2: Recognition rates (%) for SIFT descriptors VS number of dimensions.

57.1% (±2.2), max 47.6% (±3.9). These are significantly lower than the results obtained by concatenation and subsequent PCA.

9.2. Interest Points

The above results were computed using the luminance (greyscale) channel for interest point location. We also experimented with using near-infrared interest points, and using the 1^{st} PCA component. The recognition rates were 69.9% (±4.4) for NIR interest points, and 71.5% (±2.8) for using the 1^{st} PCA component. The best result was obtained using greyscale interest points 73.1% (±3.3), although the differences are within experimental error.

Colour	Descriptor Algorithm		
	HMAX	GIST	SIFT
l	70.6 (±4.6)	76.9 (±3.9)	68.0 (±3.2)
rgb	**74.0** (±4.4)	76.0 (±3.8)	67.9 (±4.0)
opp	69.6 (±3.3)	**77.8** (±3.4)	**69.0** (±4.3)
nopp	70.6 (±4.0)	75.3 (±3.3)	**69.6** (±2.5)
lrg	72.3 (±4.3)	76.6 (±3.6)	65.3 (±2.9)
pca1	70.5 (±4.7)	**77.0** (±3.8)	68.3 (±3.0)
pca2	**73.0** (±5.9)	76.8 (±4.4)	**69.6** (±3.2)
pca3	**72.8** (±5.2)	**77.3** (±4.4)	68.0 (±4.4)

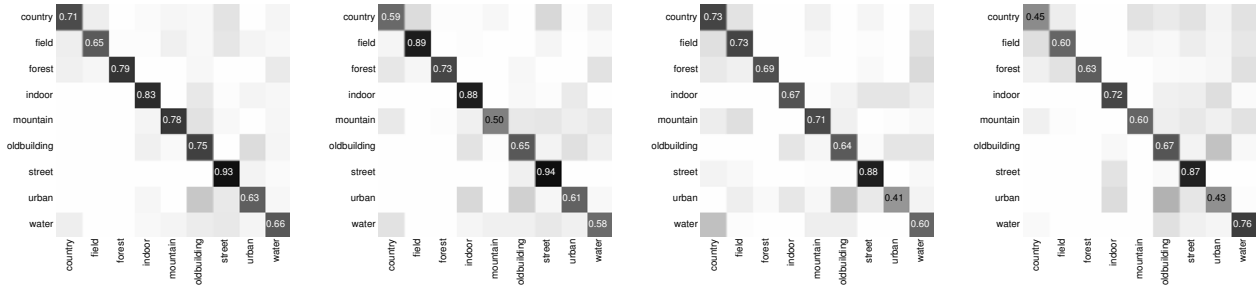
Table 3: Recognition rates (%) for Torralba's dataset.

9.3. Confusions

In general, confusions (Figure 9) are predictable from the classes, e.g., Old Buildings are often confused with Urban, and Country is often confused with Field and Forest. In going from greyscale to MSIFT the most dramatic increases are for country (45% to 71% correct), mountain (60% to 78%) and urban (43% to 63%).

10. Results on Torralba's Dataset

We also tested RGB colour only versions of the above methods on Torralba's scene categorisation dataset. This consists of 2688 images in 8 categories. To speed computation and allow us to compute statistical performance measures, we used smaller subsets for testing, consisting of 600 training and 120 test images. Again, we repeated classification experiments 10 times, using training and test images selected randomly from the whole 2688 image set. The re-



(a) msift (b) opp_sift (c) li_sift (d) sift

Figure 9: Confusion tables for scene recognition with multispectral SIFT descriptors, using various colour transformations

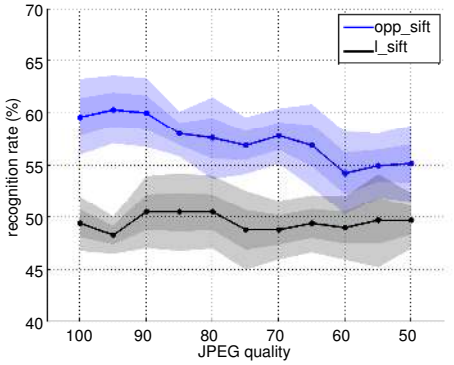


Figure 10: Recognition performance of SIFT and opponent-SIFT descriptors as the JPEG compression is increased. Note that the opponent colour descriptors are more adversely affected than their greyscale counterparts.

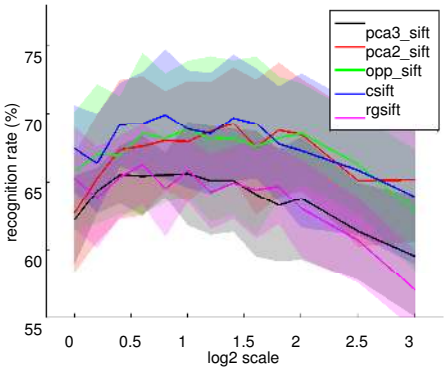


Figure 11: Chroma scale selection. Sampling the chrominance parts of colour SIFT descriptors at a lower frequency than the luminance gave improved results. We found the optimum chrominance sampling scale to be around $2\times$ the luminance sampling scale (i.e., $\log_2$ scale = 1).

sults are shown in Table 3.

One issue with Torralba's dataset is tht the images are JPEG compressed. We found that this hurt the performance of the opponent colour based descriptors, as JPEG compression introduces significant artifacts in the colour bands (e.g., chrominance information is encoded at 50% resolution). Figure 10 shows the performance of opponent SIFT descriptors as the JPEG compression is increased. Note that the greyscale SIFT descriptors suffer no loss in performance, whilst the colour descriptors' performance is significantly degraded as the JPEG compression increases.

To counteract this issue, we experimented with using descriptors where the chromatic elements were sampled at a lower frequency than the luminance parts, so that they would be less affected by compression artifacts. The results are shown in Figure 11. In all cases, better performance was achieved by increasing the sampling scale for the chromatic elements by a factor of 2 compared to the luminance sampling.

A further feature of Torralba's dataset is that, although it is highly regular in terms of scene shape, the scene colours vary tremendously. Since the the opponent algorithms (opp, nopp, lrg, pca2, pca3) are not invariant to changes in illumination colour, we experimented with colour constant versions that pre-normalise the R, G, B channels to a canonical average value of mid-grey. This gave further small performance improvements, e.g., 65.6% to 68.0% in the case of **pca3_sift**.

Overall GIST (**opp_gist**, 77.8%) performed best on this dataset, although HMAX (**rgb_hmax**, 74.0%) also gave good performance. The superiority of GIST and HMAX over bag-of-descriptors methods might be expected from the high degree of spatial regularity in the dataset. However, the results show a much weaker dependence on colour, with opponent and pca colour transformations giving only small improvements. Initial results suggest that further experimentation with colour-constant and chroma-subsampled descriptors would be worthwhile.

11. Conclusion

We explored the idea that near-infrared (NIR) information, captured from an ordinary digital camera, could be useful in visual recognition. We showed that NIR has significantly lower correlations with RGB than they do to each other, and that multispectral SIFT descriptors (MSIFT) can effectively exploit these differences to obtain improved scene recognition performance. We tested our new algorithms using a new dataset of 477 colour and near-infrared image pairs, showing significantly better performance for MSIFT than colour SIFT, HMAX and GIST. We also performed testing of colour only variants on Torralba's dataset, suggesting improvements to reduce dimensionality and increase recognition performance of colour SIFT.

12. Acknowledgements

This work is supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center supported by the Swiss National Science Foundation under grant number 5005-67322, and Xerox Foundation.

References

[1] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *International Conference on Computer Vision and Pattern Recognition*, June 2008.

[2] G. Buchsbaum and A. Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal Society*, B(220):89–113, 1983.

[3] G. Burghouts and J. Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113:48–62, 2009.

[4] L. Fei-Fei, R. VanRullen, C. Koch, and P. Perona. Why does natural scene categorization require little attention? Exploring attentional requirements for natural and synthetic stimuli. *Visual Cognition*, 12(6):893–924, 2005.

[5] D. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5:5–36, 1990.

[6] B. Funt and M. Drew. Color constancy computation in near-mondrian scenes using a finite dimensional linear model. In *International Conference on Computer Vision and Pattern Recognition*, page 544, 1988.

[7] J. Geusebroek, R. van den Boomgaard, A. Smeulders, and H. Geerts. Color invariance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(12), December 2001.

[8] V. Goffaux, C. Jacques, A. Mouraux, A. Oliva, and P. Schyns. Diagnostic colours contribute to the early stages of scene categorization: Behavioural and neurophysiological evidence. *Visual Cognition*, 12:878–892.

[9] G. Heitz and D. Koller. Learning spatial context: Using stuff to find things. In *European Conference on Computer Vision*, Marseille, 2008.

[10] D. Krishnan and R. Fergus. Dark flash photography. *ACM Transactions on Graphics, SIGGRAPH 2009 Conference Proceedings*, 2009.

[11] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[12] Y. Lu, C. Fredembach, M. Vetterli, and S. Süsstrunk. Designing color filter arrays for the joint capture of visible and near-infrared images. In *IEEE International Conference on Image Processing*, Cairo, November 2009.

[13] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *International Conference on Computer Vision*, volume 2, pages 1792–1799, 2005.

[14] N. Morris, S. Avidan, W. Matusik, and H. Pfister. Statistics of infrared images. In *International Conference on Computer Vision and Pattern Recognition*, July 2007.

[15] J. Mutch, U. Knoblich, and T. Poggio. CNS: a GPU-based framework for simulating cortically-organized networks. Technical Report MIT-CSAIL-TR-2010-013 / CBCL-286, Cambridge, MA, February 2010.

[16] J. Mutch and D. G. Lowe. Object class recognition and localization using sparse features with limited receptive fields. *International Journal of Computer Vision*, 80(1):45–57, October 2008.

[17] A. Oliva and P. Schyns. Diagnostic colors mediate scene recognition. *Cognitive Psychology*, 41:176–210, 2000.

[18] A. Oliva and A. Torralba. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 42(3):145–175, 2001.

[19] N. Pinto, D. Cox, and J. DiCarlo. Why is real-world visual object recognition hard? *PLoS Computational Biology*, 2008.

[20] D. Ruderman, T. Cronin, and C. Chiao. Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America*, 15(8):2036–2045, August 1998.

[21] N. Salamati, C. Fredembach, and S. Süsstrunk. Material classification using color and nir images. In *IS&T/SID 17th Color Imaging Conference*, 2009.

[22] L. Schaul, C. Fredembach, and S. Süsstrunk. Color image dehazing using the near-infrared. In *IEEE International Conference on Image Processing*, 2009.

[23] P. Schyns and A. Oliva. From blobs to boundary edges: Evidence for time and spatial-scale dependent scene recognition. *Psychological Science*, 5(4):195–200, 1994.

[24] R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Computer Vision*, 2(1):1–104, December 2006.

[25] A. Torralba. Contextual priming for object detection. *International Journal of Computer Vision*, 53(2):169–191, 2003.

[26] K. van de Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1582–1596, 2010.

[27] L. Zhang, B. Wu, and R. Nevatia. Pedestrian detection in infrared images based on local shape features. In *International Conference on Computer Vision and Pattern Recognition*, Minneapolis, June 2007.