On the Use of Feature Descriptors on Raw Image Data

Alina Trifan and António J. R. Neves
Universidade de Aveiro, IEETA/DETI - IRIS Laboratory, Aveiro, Portugal

Keywords: Bayer Pattern, Raw Data, Feature Detectors and Descriptors, Image Processing.

Abstract: Local feature descriptors and detectors have been widely used in computer vision in the last years for solving object detection and recognition tasks. Research efforts have been focused on reducing the complexity of these descriptors and improving their accuracy. However, these descriptors have not been tested until now on raw image data. This paper presents a study on the use of two of the most known and used feature descriptors, SURF and SIFT, directly on raw CFA images acquired by a digital camera. We are interested in understanding if the number and quality of the keypoints obtained from a raw image are comparable to the ones obtained in the grayscale images, which are normally used by these transforms. The results that we present show that the number and positions of the keypoints obtained from grayscale images are similar to the ones obtained from CFA images and furthermore to the ones obtained from grayscale images that resulted directly from the interpolation of a CFA image.

1 INTRODUCTION

Feature extraction has been studied intensively in the last years within the computer vision community. With the introduction of algorithms like Speeded Up Robust Features - SURF (H. Bay et al., 2008) and Scale Invariant Feature Transform - SIFT (Lowe, 2004), generic objects at different scales or orientations can be successfully detected in images. These two algorithms are considered to be optimal but the price paid for a good object detection is a computational time of the orders of minutes.

Most modern digital cameras allow the acquisition of images as raw data, that have a pixel distribution following the Bayer pattern (Bayer, 1976). A Bayer filter mosaic is a type of Color Filter Array (CFA) for arranging RGB color filters on a square grid of photosensors. The filter pattern is 50% green, 25% red and 25% blue, usually called BGGR, RGBG, GRGB, RGGB, etc. depending on the position of the filters.

For display purposes and better human visualization, interpolating or demosaicing algorithms are used, that convert the raw image to a certain color space, like RGB, YUV or HSV (Kimmel, 1999). This is a digital image processing technique used to reconstruct a full color image from the incomplete color samples output from an image sensor overlaid with a CFA. Most modern digital cameras acquire images using a single image sensor overlaid with a CFA, so demosaicing is part of the processing pipeline required to render these images into a viewable format. However, in most of them, images in a raw format can be retrieved, allowing the user to demosaic them using software, rather than using the cameras built-in firmware.

A recent study (Neves and Trifan, 2015) shows that there are several advantages in using CFA images for colored object detection, mainly in what concerns speeding up the processing time and the reduction of the delay between perception and action. This is due to two reasons: dealing with a reduced volume of data (a single channel image instead of a three channel one) improves the speed of the image transmission between the camera and the computer; less time is taken by the processing pipeline since demosaicing is not performed. Moreover, it has been shown in the same study that the performance of colored object detection algorithms is not affected when processing directly the CFA images.

In this paper we are interested in finding out if the same applies for algorithms used for generic object detection. By using feature descriptors and detectors such as SURF or SIFT directly on the raw CFA images, the demosaicing step is not necessary. This reduces the complexity of an object detection system. Feature descriptors and detectors such as SURF or SIFT are normally applied to intensity images, which are single channel grayscale images obtained from the full RGB images by applying a transformation that relates the intensity of a pixel with its color. In this

Trifan, A. and Neves, A.
On the Use of Feature Descriptors on Raw Image Data.
DOI: 10.5220/0005756506550662
In *Proceedings of the 5th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2016)*, pages 655-662
ISBN: 978-989-758-173-1

paper we quantify and qualify the keypoints obtained using the two mentioned feature descriptors when applied directly on CFA images and, for comparison, when applied to a grayscale image which has been obtained directly from a CFA one, based on the algorithm presented in Section 3.

The experimental results presented in this paper show that the number of keypoints obtained in the two types of images referred above are similar to the ones obtained in the intensity images and are located in similar positions. Comparing the obtained descriptors for each keypoint using the FLANN algorithm, we noticed that there are a considerable amount of them that have a match in the intensity image, mainly in the regions of the image with more detail, as desirable. We conclude that feature descriptors and detectors can be used with success directly in CFA images. As far as we know, no previous study on this matter has been presented before.

This paper is structured in 5 sections, first of them being this Introduction. An overview of the feature descriptors used in this study is presented in Section 2. Section 3 details the particularities of a Color Filter Array (CFA) image and presents the methods used for obtaining an intensity grayscale image from a CFA image. Experimental results of the use of the SIFT and SURF detectors on grayscale images, CFA images and grayscale images obtained directly from CFA images are presented in Section 4. Finally, section 5 draws the final remarks, followed by the acknowledgement of the institutions that supported this work.

2 FEATURE DESCRIPTORS AND DETECTORS

Scale-invariant feature transform (Lowe, 2004) is a popular algorithm for the detection and description of local features in an image. The SIFT descriptor is invariant to translations, rotations and scaling transformations in the image domain. It is robust to moderate perspective transformations and illumination variations. The SIFT algorithm operates in a stack of gray-scale images with increasing blur, obtained by the convolution of the initial image with a variable-scale Gaussian. A differential operator is applied in the scale-space, and candidate keypoints are obtained by extracting the extrema of this differential.

A SIFT keypoint is a selected image region with an associated descriptor. Their descriptors are stored in a vector that contains the information necessary to classify a keypoint. It is possible to obtain features, highly distinctive, useful in the matching process. In order to achieve rotation invariance, each keypoint is assigned a magnitude and an orientation, thus making this algorithm highly robust.

Speeded Up Robust Feature (H. Bay et al., 2008) is a fast and robust algorithm for local, similarity invariant representation and comparison. Similarly to the SIFT approach, SURF is a detector and descriptor of local scale and rotation-invariant image features. The SURF method uses integral images in the convolution process, which speeds up the processing. Initial images are convolved with box filters at several different discrete size. To select interest point candidates, the local maxima of a Hessian matrix is computed and a quadratic interpolation is used to refine the location of candidate keypoints. Contrast signs of the interest point are stored to construct the keypoint descriptor. Finally, the dominant orientation of each keypoint is estimated and the descriptor is computed.

SURF keypoints are assigned a scale and a rotation invariance in order to achieve distinctive features in an image. The SURF descriptor is an improvement of SIFT with respect to the processing time taken. Integral images associated with the Laplacian of Gaussian approximation represent an ingenious construction to speed up the convolution operation.

Features from Accelerated Segment Test - FAST (Rosten and Drummond, 2006) is a more recent algorithm proposed originally for identifying corners in an image. This algorithm is an attempt to solve a common problem, the one of real-time processing, with applications in robotics. Unlike SIFT and SURF, FAST algorithm only detects corners/keypoints and does not produce descriptors. This detector can be used with other descriptors to detect keypoints.

The BRIEF (Calonder et al., 2010) algorithm was the first binary descriptor published, based on simple intensity difference tests. BRIEF takes only the information at single pixels location to build the descriptor. In order to improve its sensitiveness to noise, the image is first smoothed by a Gaussian filter. This is done by picking pairs of pixels around the keypoint, according to a random or non-random sampling pattern, and then comparing the two intensities.

Although these algorithms are of great interest within the Computer Vision research community, their use has not been tested so far on raw image data. A very recent work (Larabi and Setitra, 2015) presents a preliminary study on their use on binary images. In this paper we provide results on the use of SURF and SIFT descriptors on CFA images, acquired by a digital camera. Nowadays modern digital cameras allow the acquisition of images as raw data, that have a pixel distribution following the Bayer pattern (Bayer, 1976). This work focuses on SURF and SIFT descriptors since even though they are more complex, they are the most reliable in terms of accuracy and invariance to scale and rotation (Miksik and Mikolajczyk, 2012).

3 BAYER IMAGES

Fig. 1 shows a typical Bayer arrangement of color filters. As it can be seen, the green information has double the size of the red or blue information. This is due as an attempt to mimic the physiology of the human eye, which is more sensitive to green light.

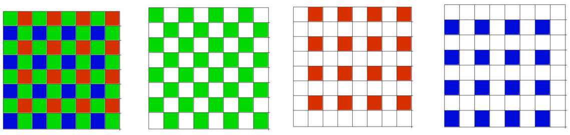

Figure 1: Bayer arrangement of color filters.

To obtain a full-color image, various demosaicing algorithms can be used to interpolate a set of complete red, green, and blue values for each pixel. These algorithms make use of the surrounding pixels of the corresponding colors to estimate the values for a particular pixel. In Fig. 2 we can see an example of a CFA image and the corresponding RGB image obtained after demosaicing.

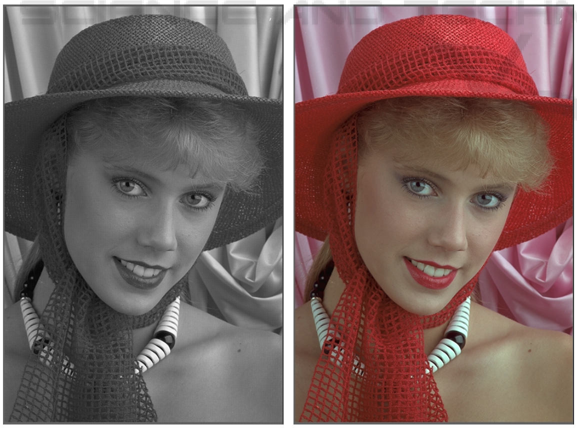

Figure 2: On the left an example of a CFA image. On the right, the obtained RGB image after demosaicing.

In this paper we provide experimental results regarding the use of feature descriptors like SIFT and SURF directly on CFA images. Moreover, for comparison, we provide experimental results using intensity grayscale images obtained directly from the CFA images. To do that, we consider the neighbor information of red, green and blue to calculate the intensity information of each pixel without any interpolation.

The algorithm for transforming a CFA image into a grayscale image, considering the Bayer configuration presented in Fig. 1, works as follows:

```
Input: CFA image (image)
Output: grayscale image (y)

/* For all the pixels in the Bayer image */
for(p = 0 ; p < image.cols * image.rows ; p++)
{
  row = p / image.cols;
  col = p \% image.cols;

  if(row \% 2 == 0) /* even rows */
  {
    if(col \% 2 == 0) /* even columns */
    {
      r = image[p];
      g = image[p + 1];
      b = image[p + image.cols + 1];

    }
    else
    {
      g = image.ptr[p];
      r = image.ptr[p - 1];
      b = image.ptr[p + image.cols];

    }
  }
  else /* odd rows */
  {
    if(col \% 2 == 0) /* even columns */
    {
      g = image.ptr[p];
      r = image.ptr[p - image.cols];
      b = image.ptr[p + 1];

    }
    else /* odd columns */
    {
  g = image.ptr[p - 1];
      r = image.ptr[p - image.cols - 1];
      b = image.ptr[p];
    }
  }
  /* Grayscale pixel */
  y[p] = 0.299 * r + 0.587 * g + 0.114 * b;
}
```

Fig. 3 shows an example of the application of the previous algorithm. Visually, the image obtained represents the light intensity on each pixel without noise. For comparison, in Fig. 4 we present the grayscale version obtained from the full RGB image and the corresponding difference image. The difference image shows the absolute differences between this image and the one obtained directly from the CFA image. The most considerable differences are in the pixels where there are edges, as expected, since in more flat regions the interpolation process usually does not introduce new color information.

Figure 3: On the left an example of a CFA image. On the right, the obtained grayscale image using directly the CFA image.

Figure 4: On the left, the grayscale image obtained from the full RGB image. On the right, the difference between the image on the left and the grayscale image obtained from the CFA image presented in Fig. 3.

4 EXPERIMENTAL RESULTS

In order to quantify and qualify the keypoints obtained using SIFT and SURF feature descriptors when applied directly on CFA images and, for comparison, when applied to a grayscale image obtained directly from the CFA ones, we used the SIFT and SURF implementation provided by the OpenCV library.

In order to perform the experiments reported in this paper, twenty-four 24-bit color images from the well known Kodak image set (kod,) of size 512×768 each, as shown in Fig. 15, were sub-sampled according to the Bayer pattern presented in Fig. 1 to form a set of 8-bit testing raw Bayer images. We obtained a set of 8-bit intensity grayscale images obtained directly from the RGB original images and another 8-bit intensity grayscale images obtained directly from the CFA images using the algorithm described in Section 3. Figure 5 shows the keypoints detected for image 04 of the used dataset, using SIFT. The results presented on the left have been obtained using the grayscale image. The keypoints of the central image have been obtained by applying the SIFT transform directly on the CFA image. The image on the right is the grayscale image that has been obtained directly from the CFA one.

Figure 5: SIFT keypoints for images 04 (first row) and 07 (sencond row) of the Kodak set. The complete results can be found in http://sweet.ua.pt/an/FullResults.zip.

Figure 6 shows the keypoints detected for image 04 of the used dataset, using SURF. The results presented on the left have been obtained using the grayscale image of the dataset. The keypoints of the central image have been obtained by applying the SURF transform directly on the CFA image. The image on the right is the grayscale image that has been obtained directly from the CFA one.

The results presented in these images show that the number and position of the keypoints are similar between gray, CFA and grayscale images obtained from the CFA ones. Moreover, Table 1 and Table 2 present detailed experimental results regarding the number of keypoints obtained using the SIFT and SURF algorithms when applied to the intensity grayscale image obtained from the full RGB image, when applied to the CFA image and when applied to the grayscale image obtained from the CFA image.

To measure the quality of the obtained descriptors, we compared the descriptors obtained for the keypoints in the CFA images and for the ones in grayscale images obtained from CFA images, with the ones in intensity grayscale images obtained from the full RGB images. This comparison was based on the FLANN matching algorithm (Muja and Lowe, 2009).

Figure 6: SURF Keypoints for for images 04 (first row) and 07 (sencond row) of the kodak set. The complete results can be found in http://sweet.ua.pt/an/FullResults.zip.

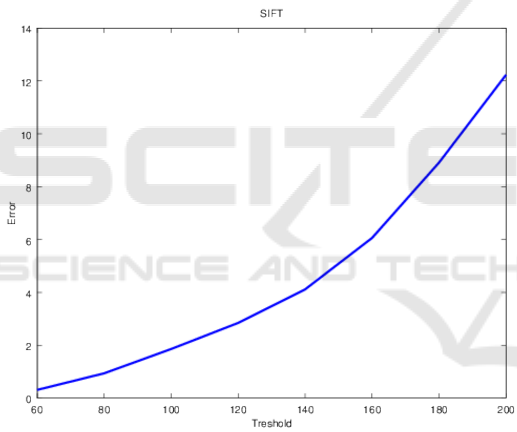

Figure 7: Average error between the position of the SIFT keypoints in the CFA image and the position of the SIFT keypoints in the grayscale image. Different threshold values of the FLANN matching algorithm have been used.

FLANN is an algorithm for performing fast approximate nearest neighbor searches in high dimensional spaces. It contains a collection of algorithms found to work best for nearest neighbor search and a system for automatically choosing the best algorithm and optimum parameters depending on the dataset.

Experimental results presented in Table 1 and Table 2 show the number of matches for each pair of images. We can observe that there is a considerable number of matches below a low error. We considered an error of 0.08 for the SURF algorithm for all the images and an error of 100 for the SIFT algorithm.

In Fig. 7 and Fig. 8 we can see the average error in the position of the keypoints in the CFA image and

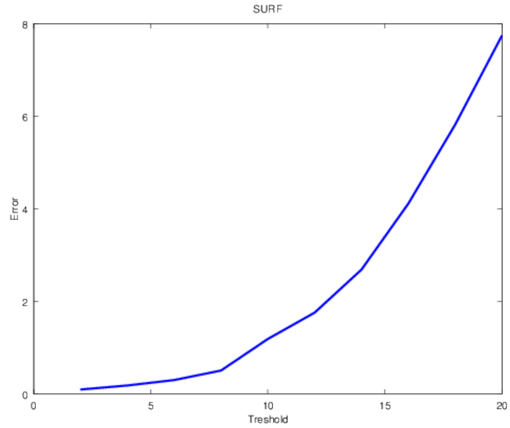

Figure 8: Average error between the position of the SURF keypoints in the CFA image and the position of the SURF keypoints in the grayscale image. Different threshold values of the FLANN matching algorithm have been used.

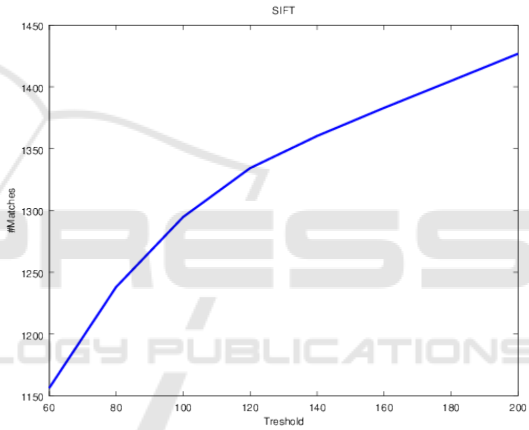

Figure 9: Number of SIFT keypoints in the CFA image that corresponds to SIFT keypoints in the grayscale image. Different threshold values of the FLANN matching algorithm have been used.

the position of the keypoints in the grayscale image for both algorithms. The average error was obtained for all the 24 images of the dataset.

The threshold for the error between matches has an effect on the number of matches, as presented in Fig. 9 and Fig. 10. If necessary in a specific application, we can improve the number of matches and the error in their position adjusting this threshold.

Figure 11 shows the matches obtained using the FLANN matcher between the SURF keypoints found in intensity image 04 of the used dataset and the corresponding CFA image. Figure 12 shows the matches obtained using the FLANN matcher between the SURF keypoints found in intensity image 04 of the used dataset and the grayscale obtained from the CFA image. It can be observed that the position of

Table 1: Table containing information about the number of keypoints obtained using the SIFT algorithm when applied to the grayscale image obtained from the full RGB image (column "#kG"), when applied to the CFA image (column "#kB") and when applied to the grascale image obtained from the CFA image (column "#kGB"). This table also shows the number of keypoints in the CFA image that corresponds to keypoints in the grayscale image using the FLANN algorithm (column "#mB") and the number keyoints in the grayscale image obtained from the CFA image that corresponds to keypoints in the grayscale image.

Img	#kG	#kB	#mB	#kGB	#mGB
1	3199	2737	2094	2915	840
2	547	277	191	411	141
3	957	647	471	668	317
4	1465	815	619	891	372
5	4052	3605	2755	3217	1374
6	2118	1725	1279	1831	576
7	1899	1506	1149	1441	821
8	3603	3274	2703	2927	1207
9	1393	1259	1018	971	489
10	1355	1246	998	968	510
11	1637	1405	1044	1324	556
12	648	574	414	450	203
13	4187	3366	2481	3569	1072
14	3012	2522	1913	2308	956
15	935	664	524	526	266
16	1307	1063	811	980	378
17	1412	1303	1023	1092	615
18	4027	2633	1999	2652	1017
19	1531	1093	826	1060	444
20	890	713	521	688	261
21	2590	2199	1709	2099	844
22	1932	1334	952	1253	469
23	727	570	419	491	266
24	3084	2611	1866	2262	839

Table 2: Table containing information about the number of keypoints obtained using the SURF algorithm when applied to the grayscale image obtained from the full RGB image (column "#kG"), when applied to the CFA image (column "#kB") and when applied to the grascale image obtained from the CFA image (column "#kGB"). This table also shows the number of keypoints in the CFA image that corresponds to keypoints in the grayscale image using the FLANN algorithm (column "#mB") and the number keyoints in the grayscale image obtained from the CFA image that corresponds to keypoints in the grayscale image.

Img	#kG	#kB	#mB	#kGB	#mGB
1	2819	3846	1117	2526	747
2	246	415	50	247	81
3	455	786	135	447	269
4	537	979	185	515	216
5	2650	3844	1175	2634	983
6	1268	1908	443	1102	357
7	1193	2440	513	1172	587
8	3336	4270	1900	2921	804
9	885	1237	538	846	415
10	788	1077	464	775	366
11	1238	1839	525	1145	476
12	493	969	284	488	256
13	2669	4134	905	2393	700
14	1850	2659	585	1760	710
15	515	863	306	569	270
16	529	900	270	474	226
17	1019	1426	616	1011	529
18	1804	2522	498	1659	535
19	1198	2406	537	1069	386
20	674	1258	356	710	313
21	1623	1258	356	1494	586
22	1000	2021	214	888	289
23	476	853	171	476	234
24	1909	2729	812	1772	613

matches are mainly in relevant parts of the images, enough to describe the object of interest.

Figure 13 shows the matches obtained using the FLANN matcher between the SIFT keypoints found in intensity image 04 of the used dataset and the corresponding CFA image. Figure 14 shows the matches obtained using the FLANN matcher between the SIFT keypoints found in intensity image 04 of the used dataset and the grayscale obtained from the CFA image. It can be observed that the position of matches are mainly in relevant parts of the images, enough to describe the object of interest.

5 CONCLUSIONS

We have presented in this paper a study on the application of two of the most used feature descriptors, SURF and SIFT, on raw CFA images. The results that we presented prove that it is possible to use these descriptors directly on CFA images, discarding thus the need of interpolating a raw image into a full RGB one prior to processing it. We have presented comparative results of the use of the two transforms on intensity grayscale images obtained from full RGB images, CFA images and intensity grayscale images obtained from raw CFA images by direct conversion. This study is an important contribution for the Computer Vision community since it proves that generic object detection can be done directly on raw images

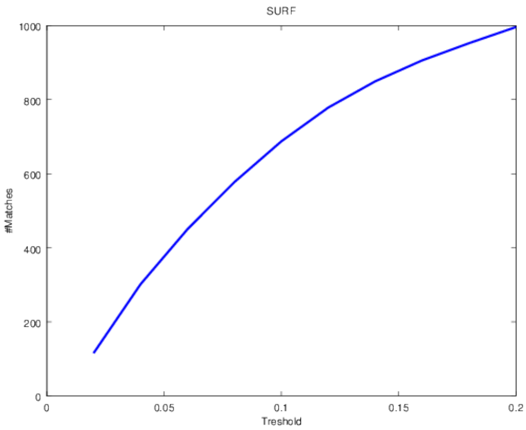

Figure 10: Number of SURF keypoints in the CFA image that corresponds to SURF keypoints in the grayscale image. Different threshold values of the FLANN matching algorithm have been used.

Figure 11: Match SURF Keypoints for image 04 of the Kodak set (on the left) using FLANN algorithm considering the CFA image (on the right). The complete results can be found in http://sweet.ua.pt/an/FullResults.zip.

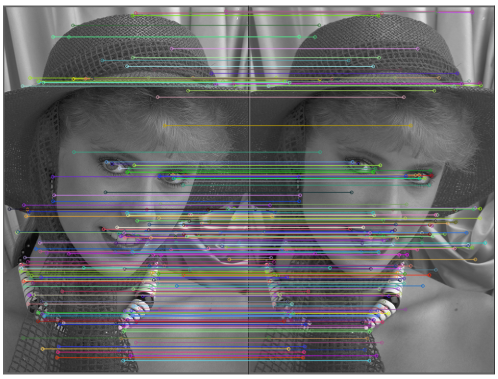

Figure 12: Match SURF Keypoints for image 04 of the Kodak set (on the left) using FLANN algorithm considering the grayscale obtained from the CFA image (on the right). The complete results can be found in http://sweet.ua.pt/an/FullResults.zip.

and the demosaicing of these images is no longer a compulsory step in an image processing pipeline.

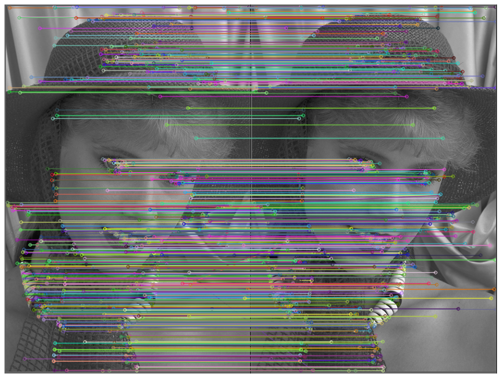

Figure 13: Match SIFT Keypoints for image 04 of the Kodak set using FLANN algorithm. The complete results can be found in http://sweet.ua.pt/an/FullResults.zip.

Figure 14: Match SIFT Keypoints for image 04 of the Kodak set using FLANN algorithm. The complete results can be found in http://sweet.ua.pt/an/FullResults.zip.

Future work directions will focus on the use of these descriptors directly on the raw data for object detection. For this, a more detailed study on the threshold and parameters involved by the matching algorithm will be conducted.

ACKNOWLEDGEMENTS

This work was developed in the Institute of Electronic and Telematic Engineering of University of Aveiro and was partially supported by FEDER through the Operational Program Competitiveness Factors - COMPETE FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011) and by National

Figure 15: Grayscale version of the Twenty-four digital color images from the kodak set (refers as image 1 to image 24, from top-to-bottom and left-to-right).

Funds through FCT - Foundation for Science and Technology in a context of a PhD Grant (FCT reference SFRH/BD/85855/2012).

REFERENCES

Kodak image set. http://r0k.us/graphics/kodak/. Accessed: 2015-10-5.

Bayer, B. (1976). Color imaging array. US Patent 3,971,065.

Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). BRIEF: Binary Robust Independent Elementary Features. *11th European Conference on Computer Vision (ECCV), Heraklion, Crete. LNCS Springer*.

H. Bay, A. E., Tuytelaars, T., and Gool, L. V. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359.

Kimmel, R. (1999). Demosaicing: image reconstruction from color ccd samples. *IMAGE PROCESSING, IEEE TRANSACTIONS ON*.

Larabi, S. and Setitra, I. (2015). A study on discrimination of sift feature applied to binary shapes. In *Proc. of ECCOMAS Thematic Conferences on Computational Vision and Medical Image Processing VIP 2015, Santa Cruz de Tenerife, Spain*, pages 295–301. Taylor and Francis.

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

Miksik, O. and Mikolajczyk, K. (2012). Evaluation of local detectors and descriptors for fast feature matching. In *ICPR*, pages 2681–2684. IEEE.

Muja, M. and Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. *International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 331–340.

Neves, A. and Trifan, A. (2015). Time-constrained detection of colored objects on raw bayer data. In *Proc. of ECCOMAS Thematic Conferences on Computational Vision and Medical Image Processing VIP 2015, Santa Cruz de Tenerife, Spain*, pages 288–294. Taylor and Francis.

Rosten, E. and Drummond, T. (2006). Machine learning for high speed corner detection. *9th European Conference on Computer Vision (ECCV)*, 1:430–443.