

SIFT IN PERCEPTION-BASED COLOR SPACE

Yan Cui Alain Pagani Didier Stricker

DFKI, Augmented Vision
Kaiserslautern University, Germany

ABSTRACT

Scale Invariant Feature Transform (SIFT) has been proven to be the most robust local invariant feature descriptor, however, SIFT is designed mainly for grayscale images. Many local features can be misclassified if their color information is ignored. Motivated by perceptual principles, this paper addresses a new color space, called perception-based color space, in which the associated metric approximates perceived distances and color displacements and capture relationships that are illumination invariance. Instead of using grayscale values to represent the input image, the proposed approach builds the SIFT descriptors in the new color space, resulting in a descriptor that is more robust than the standard SIFT with respect to color and illumination variations. The evaluation results support the potential of the proposed approach.

Index Terms— SIFT, color space, local features

1. INTRODUCTION

In all of the feature extraction methods, the invariance with respect to imaging conditions represents the biggest challenge. More specifically, the local extracted features should be invariant with respect to geometrical variations, such as translation, rotation, scaling, and affine transformations. Furthermore, these features should be invariant with respect to photometric variations such as illumination direction, intensity, colors, and highlights. SIFT [1] [2] has been proven to be the most robust among the other local invariant feature descriptors with respect to different geometrical changes [3]. However, due to the color constancy problem, a lot of geometrical invariant approaches avoid dealing with colored images. Therefore, illumination invariance is a crucial problem which has to be solved for local features. While some researchers already focused on the color constancy problem [4] [5], some attempts to make use of the color information inside the SIFT descriptors have been proposed [6] [7] [8]. In [6], the normalized RGB model has been used in combination with SIFT to achieve partial illumination invariance besides its geometrical invariance. The color invariance of this approach is still limited because of the primitive color model used. In [7], a multi-stage recognition approach has been developed in order to achieve both color and geometrical invariance. In the first

stage, a color classifier is used label the different image regions. Then, the SIFT descriptors are augmented by adding the color labels. In spite of the good performance of this approach, its need for colored learning instances limits its performance in several applications. In [8], physical-based color invariants have been developed for invariant color representations under different imaging conditions. This color invariance model, called CSIFT, results from the Kubelka-Munk theory [9], and parameters have to be provided by the user.

The model we propose in this paper is called Perception-based Color SIFT (PC-SIFT). It provides geometrical and photometric invariance. PC-SIFT is based on standard SIFT method: the locality of the extracted features and the way in which the descriptors are built provides the invariance with respect to the geometrical variations, and the scale-space theory offers the main tool for selecting the most robust feature locations against scale variations. At the same time, PC-SIFT is based on the perception-based color space [10], which provides invariance to illumination. Firstly, we seek a color space where difference vectors between color pixels are unchanged by reillumination. We therefore restrict our attention to 3-dimensional color space parameterizations in which color displacements, or gradients, can be computed simply as component-wise subtractions. Secondly, the l_2 norm of a difference vector should match the perceptual distance between the two colors. Therefore, the standard computational method of measuring error and distance in color space should match the perceptual metric used by human viewers. This principle relates to the idea of “flatness” of perceptual space, where perceptual distance can be computed by the Euclidean distance when the Euclidean computation is preceded by an appropriate nonlinear reparameterization of color space (as in color spaces such as CIE $L^*a^*b^*$ and CIE $L^*u^*v^*$). There is ample evidence that perceptual space is unlikely to be exactly flat [11], but there is also a fair amount of evidence that it may be usefully treated as flat for many applications [11] [12].

The remainder of the paper is organized as follows: We will present the conversion from RGB color space to the perception-based color space in more details in the section 2. In section 3 the SIFT method in the perception-based color space will be explained. The evaluation results showing the high performance of PC-SIFT in comparison with the standard SIFT and CSIFT [8] are presented in the section 4. We

finally conclude in section 5.

2. PERCEPTION-BASED COLOR SPACE

In this section, we explain how to formalize the perception color space conditions [10]. Firstly, we translate RGB color space to XYZ color space. The standardized transformation settled by the CIE special commission [13], γ is a gamma corrected function with $\gamma = 2.0$, as follows:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.177} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.177 & 0.812 & 0.011 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} \gamma(R) \\ \gamma(G) \\ \gamma(B) \end{bmatrix} \quad (1)$$

Secondly, we translate XYZ color space to the perception color space UVW . Let's assume that $\vec{x}$ is the tristimulus values of a sensor represented in XYZ coordinates, F is the 3D color space parameterization we wish to solve for. Following [10], we assume that materials and illuminants in our scenes are such that the effect of relighting is well-approximated by multiplying each tristimulus value (in an appropriate basis) by a scale factor that does not depend on the materials observed. We represent the fixed change to an appropriate basis by the matrix B . Therefore, the effect of relighting can be written as

$$\vec{x} \mapsto F(\vec{x}) = B^{-1}DB\vec{x} \quad (2)$$

where D is a diagonal matrix depending only on the illuminants and not the material of objects. It's shown ([10]) that the nonlinear function F of Equation (2) must take the form:

$$F(\vec{x}) = A \left(\hat{\ln}(B\vec{x}) \right) \quad (3)$$

where A and B are invertible 3×3 matrices and $\hat{\ln}$ denotes the component-wise natural logarithm. The matrix B transforms color coordinates to the basis in which relighting best corresponds to multiplication by a diagonal matrix, while the matrix A provides degrees of freedom that can be used to match perceptual distances. In [10], the matrix A and B have been experimentally estimated using databases of similar colors. We take the same obtained values as follows:

$$A = \begin{bmatrix} 27.07439 & -22.80783 & -1.806681 \\ -5.646736 & -7.722125 & 12.86503 \\ -4.163133 & -4.579428 & -4.576049 \end{bmatrix} \quad (4)$$

$$B = \begin{bmatrix} 0.9465229 & 0.2946927 & -0.1313419 \\ -0.117917 & 0.9929960 & 0.007371554 \\ 0.0923046 & -0.04645794 & 0.9946464 \end{bmatrix} \quad (5)$$

3. PC-SIFT DESCRIPTORS

The three main stages in an invariant feature extraction method are interest points detection, descriptor building and descriptor matching.

Interest points should be selected so that they achieve the maximum possible repeatability under different photometric and geometric conditions. As discussed in section 2, our model PC-SIFT is based on the perception-based color space, which is invariant to illumination changes. Furthermore, the extremum in Laplacian pyramid, which is approximated by the Difference-of-Gaussian for the input image in different scales, has been proven to be the most robust interest points detector with respect to geometrical changes [3] [6]. We follow the same strategy as the standard SIFT method in PC-SIFT, but use our perception based color space, and detect the feature points for the three new color channels respectively.

After localizing the interest points, feature descriptors are built to characterize these points. These descriptors should contain distinct and specific information for their corresponding interest points. Different schemes have been followed for building descriptors [1] [2] [3]. Instead of using grayscale gradients for building feature descriptors, we set up a new feature descriptor in 3D perception-based color space. As Fig.1 shows, for a given pixel in the vicinity of the interest point, we denote G_x the 3-dimensional color gradient in x direction, and G_y the 3-dimensional color gradient in y direction. We consider the angle θ from vector G_x to vector G_y as the gradient orientation. The range of θ is from 0 to 2π . All orientations are assigned relative to a dominant/canonical orientation of the interest point. We consider the length of the vector from G_x to G_y as the magnitude of our local gradient. Thus, if the triangle OG_xG_y is same or similar to the triangle $OG'_xG'_y$, they are considered as the same feature. The PC-SIFT descriptor is invariant to the illumination or color transformation. As in the standard SIFT descriptor, the PC-SIFT descriptor is built as histograms with 128 dimensions.

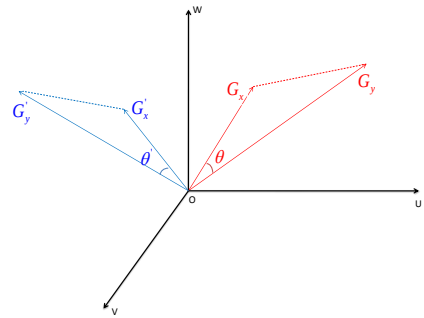


Fig. 1. The new color space and PC-SIFT descriptor .

Finally, the matching process is performed for the built local descriptors by finding the nearest neighbor for each pair of feature positions. After rejecting outliers, we can estimate the correspondences between two images, or the object pose, when the geometry is known.

4. EXPERIMENTAL RESULTS

To evaluate the proposed approach, we use the “Amsterdam Library of Object Images (ALOI)” [14] which is an image database of colored objects. ALOI contains a large number of objects under different imaging conditions, namely, different illumination directions, illumination intensities, illumination colors, and object viewpoints.

We give in Fig.2 (a) and (b) visual results under different illumination color and illumination directions and intensities with our PC-SIFT descriptor. The number of total matches and correct matches with PC-SIFT, CSIFT [8] and standard SIFT under different illumination color and illumination directions and intensities are presented in (c) and (d). There’re 12 different color temperature cases and 8 different illumination sources settings for one test object, but it’s unchanged of the object position. Our model PC-SIFT has a highest performance, because PC-SIFT can detect more feature points under the new color space. We can observe that the distance of the lines between total matches and correct matches, the PC-SIFT distance is the smallest, express that the PC-SIFT descriptor is most stable one. In Fig.2 (e) and (f), the recall precision of PC-SIFT, CSIFT and standard SIFT as a function of the total number of matches is given, where recall precision is the ratio between the number of correct matches and number of possible matches. The total number of matches can be varied by changing the threshold for the maximum allowed distance between two descriptors. Our algorithm consistently performs better than the other approaches on all test images. We also analyze the performance on the database, as Fig.2 (g) and (h) shows, we select 200 images, and the number of the correct matches with the three methods are shown, sorted by the correct mean matches number of the three methods increasing. For most cases, our method PC-SIFT is better than the other two descriptors. There’re 12 cases that CSIFT is better than PC-SIFT among 400 cases under different illumination conditions. We can conclude that the proposed approach performs better than standard SIFT and the CSIFT method.

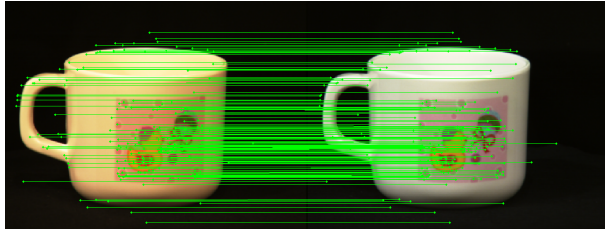
5. CONCLUSION

In this paper, we introduce PC-SIFT as a novel color-based local invariant feature descriptor for the purpose of combining both color and geometrical information in image matching. We achieved the color and illumination invariance by using a perception-based color model [10], and the geometrical invariance is achieved by building PC-SIFT using a structure similar to that of the SIFT descriptors. Evaluation results showed that PC-SIFT is superior to standard SIFT and CSIFT for color images under illumination changes and proved the high performance of PC-SIFT.

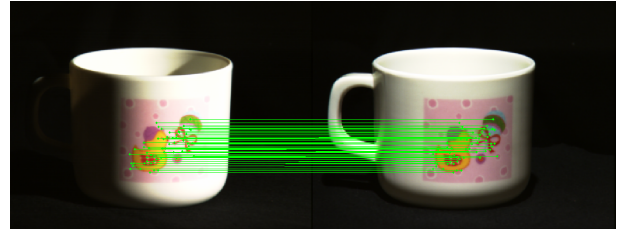
Acknowledgment - This work has been partially funded by the project CAPTURE and the German BMBF project AVILUSplus (01M08002)

6. REFERENCES

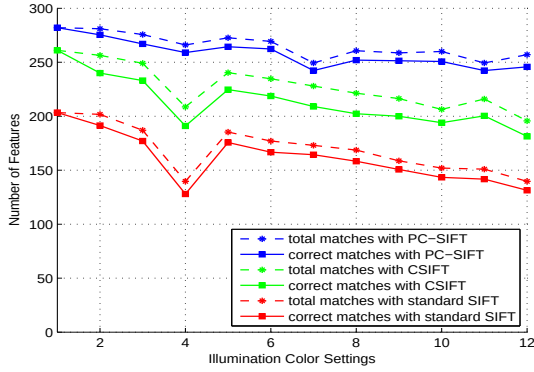
- [1] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, 1999, vol. 2, pp. 1150–1157 vol.2.
- [2] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [3] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *PAMI, IEEE Transactions on*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
- [4] D. Brainard and W. Freeman, “Bayesian color constancy,” *the Journal of Optical Society of America*, vol. 14, pp. 1393–1411, 1997.
- [5] M.D’Zmura and P. Lennie, “Mechanisms of color constancy,” *the Journal of Optical Society of America*, vol. 3, pp. 1662–1672, 1986.
- [6] M. Brown and D. G. Lowe, “Invariant features from interest point groups,” in *In British Machine Vision Conference*, 2002, pp. 656–665.
- [7] A.A. Farag and Alaa E. Abdel-Hakim, “Detection, categorization and recognition of road signs for autonomous navigation,” *Proc. of Advanced Concepts in Intelligent Vision Systems (ACIVS2004)*, pp. 125–130, 2004.
- [8] A. E. Abdel-Hakim and A. A. Farag, “Csift: A sift descriptor with color invariant characteristics,” in *CVPR, 2006 IEEE*, 2006, vol. 2, pp. 1978–1983.
- [9] P. Kubelka, “New contribution to the optics of intensely lightscattering materials,” *the Journal of Optical Society of America*, vol. 38, pp. 448457, 1948.
- [10] Y. Chong Hamilton, Steven J. Gortler, and Todd Zickler, “A perception-based color space for illumination-invariant image processing,” *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–7, 2008.
- [11] G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae (Wiley Series in Pure and Applied Optics)*, Wiley-Interscience, 2 edition, August 2000.
- [12] D. B. Judd and G. Wyszecki, *Color In Business, Science, and Industry*, John Wiley and Sons, 1975.
- [13] M. H. Brill, *How the CIE 1931 color-matching functions were derived from Wright-Guild data*, Sarnoff Corp, 1998.
- [14] G. J. Burghouts J. M. Geusebroek and A.W. M. Smeulders., “The amsterdam library of object images,” *Int. J. Comput Vision*, vol. 61, pp. 103112, 2005.



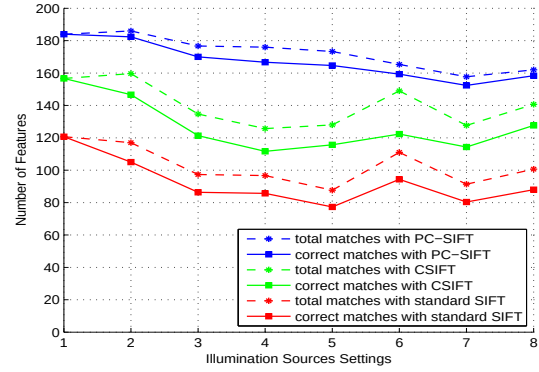
(a) Result under different illumination color



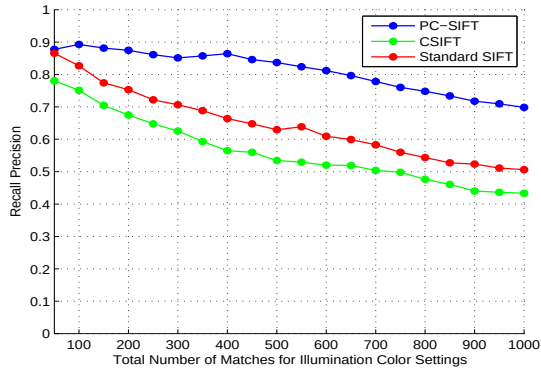
(b) Result under different illumination directions and intensities



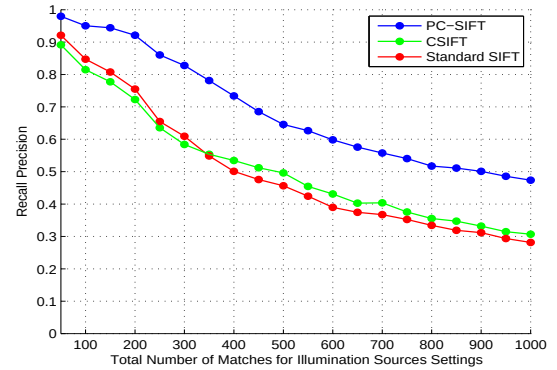
(c) Number of the detected and matched keys under different illumination colors



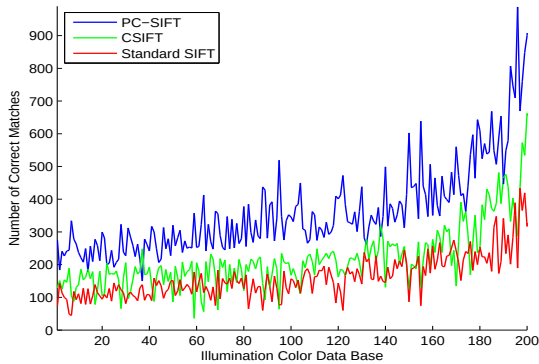
(d) Number of the detected and matched keys under different illumination directions and intensities



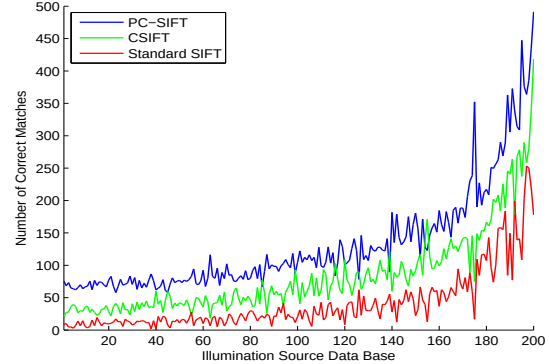
(e) Recall precision of detected features as a function of the total number of matches under different illumination colors



(f) Recall precision of detected features as a function of the total number of matches under different illumination directions and intensities



(g) Number of correct matches for the illumination colors data base



(h) Number of correct matches for the illumination directions and intensities data base

Fig. 2. Evaluation results of matching under varying illumination conditions compared with PC-SIFT, CSIFT and SIFT