2010 International Conference on Pattern Recognition

Improving SIFT-based descriptors stability to rotations

Fabio Bellavia, Domenico Tegolo
Dipartimento di Matematica e Informatica
Universitá di Palermo, 90123 Palermo, Italy
{fbellavia,domenico.tegolo}@unipa.it

Emanuele Trucco
School of Computing
University of Dundee, Dundee DD1 4HN, Scotland
manueltrucco@computing.dundee.ac.uk

Abstract

Image descriptors are widely adopted structures to match image features. SIFT-based descriptors are collections of gradient orientation histograms computed on different feature regions, commonly divided by using a regular Cartesian grid or a log-polar grid.

In order to achieve rotation invariance, feature patches have to be generally rotated in the direction of the dominant gradient orientation. In this paper we present a modification of the GLOH descriptor, a SIFT-based descriptor based on a log-polar grid, which avoids to rotate the feature patch before computing the descriptor since predefined discrete orientations can be easily derived by shifting the descriptor vector. The proposed descriptors, called sGLOH and sGLOH+, have been compared with the SIFT descriptor on the Oxford image dataset, with good results which point out its robustness and stability.

1. Introduction

Matching between different images is one of the most challenging tasks in computer vision. This is a necessary step in many computer vision applications such as three–dimensional reconstruction, mosaicing and object recognition [6, 10]. Image patches have to be extracted by a feature detector, or the whole image is used in the case of dense matching [16]. The patches are then trasformed to feature vectors by a feature descriptor algorithm, and the vectors eventually compared to establish the matches.

Different feature detectors as well as feature descriptors have been developed in the last decade. State-of-the-art detectors include corner detectors based on the autocorrelation matrix [12, 5] or blob-like detectors [12, 8, 2, 10, 11]. Although a lot of progress has been done, different evaluation tests [14] suggest that no feature detector outperforms the others substantially and results should be improved in order to obtain more stable features, especially for non-planar scenes.

In general, in order to obtain the feature vector, a support region around the extracted keypoint is computed [13] by normalizing the feature patch, to take into account geometrical transformations, luminosity changes and rotations. The normalized region is then subsampled, obtaining a feature patch or fingerprint. Affine geometric transformations are used for affine covariant feature detectors, which map ellipses to circles. To compensate for illumination changes, the region intensity values are normalized by their mean and standard deviation.

Rotation invariance is achieved with a rotation of the patch in the direction of the dominant gradient orientation. The most common approach was proposed by Lowe [10]. The gradient orientation histogram for the whole patch, weigthed by the gradient magnitude on a Gaussian window centered in the keypoint, is computed, and the dominant orientation is given by the bin with the maximum value. It should be noted that the extracted direction can be ambiguous, especially for highly symmetric patches.

The feature descriptor vector for each patch is then computed in order to match corresponding points. A widely used class of descriptors is given by the histogram–based descriptors. One of the most reliable, efficient and robust feature descriptors is the SIFT (Scale Invariant Feature Transform) descriptor [10], given by a three-dimensional histogram of gradient

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.845

locations and orientations. The SIFT descriptor has been extended in several ways, for example, the PCA–Sift [7] decrease the vector dimensionality by the Principal Component Analysis, the GLOH (Gradient Location and Orientation Histogram) [13] descriptor uses a polar grid while the IG (Irregular Grid) [4] descriptor overlaps the bins. Similar to the SIFT descriptor are the SURF descriptor [2], employing wavelets, and the DAISY descriptor [16], which has been used for dense matching.

Recently, new distance measures have been also developed to take into account the spatial relations between the histogram bins for SIFT-based descriptors, as the Diffusion Distance [9], to which also the SIFT-Rank descriptor [15], which considers the rank-order of the SIFT descriptor vector, can be associated.

In this paper we present a modification of the GLOH descriptor, which avoids rotating the feature patch before computing the descriptor since predefined discrete orientations can be easily derived by shifting the descriptor vector. Instead of rotating the patch in the estimated dominant orientation, for which a accurate computation can be difficult, the descriptor is compared with the different discrete orientations which can be obtained by shifting the vector for a reasonable computational cost, and the best is selected.

The proposed descriptors, called sGLOH and sGLOH+, have been compared with the SIFT descriptor on the Oxford image dataset [1]. Good results have beeen obtained that show the robustness and the stability of the new descriptor. In Sec. 2 the new descriptor with implementation details is given, while in Sec. 3 performances are evaluated. Lastly, in Sec. 4 conclusion and final remarks are discussed.

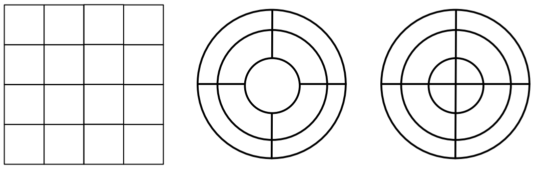

Figure 1. SIFT grid (left), sGLOH grid with $n = 2$, $m = 4$ and respectively $\Psi(H) = 0$ (center) and $\Psi(H) = 1$ (right).

2. The sGLOH descriptor

The sGLOH descriptor is based on the GLOH [13]. The descriptor grid is made up of n circular rings centered on the feature point. Each ring contains m regions, equally distributed along the m directions, defining a region $R_{r,d}$ with $r = \{1, 2, \ldots, n\}$ and $d = \{0, 1, \ldots, m-1\}$. The inner circular region can be divided in m radial sectors, as shown in Fig. 1, defining the single region $R_{0,0}$ or more regions $R_{0,d}$. Given a descriptor vector H, a function $\Psi(H)$ is defined as equal to 0 if a single region is defined, 1 otherwise.

For each region, the orientation histogram weighted by the gradient magnitude is computed in m quantized orientation. In order to obtain a better gradient distribution estimation, for each region the bin value h_i where $i = 0, 1, \ldots, m-1$ is computed with a kernel density estimation by a Gaussian window

$$h^i_{r,d} = \frac{1}{\sqrt{2\pi}\sigma} \sum_{p \in R_{r,d}} G_m(p)\, e^{-\frac{(M_{2\pi}(G_d(p)-m_i))^2}{2\sigma^2}}$$

where $G_m(p), G_d(p)$ are the gradient magnitude and orientation of a pixel p in the region $R_{r,d}$, with $r = \{0, 1, \ldots, n\}$ and $d = \{0, 1, \ldots, m-1\}$, $m_i = \frac{2\pi}{m} i$ is the i-th bin center, and $\sigma = \frac{2\pi}{m} c$, with $c \in \mathbb{R}^+$, is the standard deviation. The function $M_q(x)$ is used to take into account a periodicity of length q

$$M_q(x) = \begin{cases} x & \text{if} \quad x < \frac{q}{2} \\ q - x & \text{otherwise} \end{cases}$$

In modular arithmetic, $[i + d]_m$ shifts cyclically by d positions the element i of a vector of size m, given the congruence modulo m relation $a \equiv b(mod\ m)$, where the congruence class is represented by $[a]_m$. Defining a block histogram

$$H_{r,d} = \bigoplus_{i=0}^{m-1} h_{r,d}^{[d+i]_m},$$

where $\bigoplus$ is the contatenation operator, so that for each block the first bin has direction d, the final descriptor vector H is obtained by concatenating histograms

$$H = \bigoplus_{i=0}^{n} \bigoplus_{j=0}^{m-1} H_{i,j}$$

where $H_{0,k}$ for $k = 1, \ldots, m-1$ is not considered if $\Psi(H) = 0$. The final descriptor length is $l = m(mn + 1 + (m-1)\Psi(H))$.

The rotation of the descriptor by a factor αk where $\alpha = \frac{2\pi}{m}$ is then given by a ciclic shift of the block histogram inside a ring

$$H_{\alpha k} = \begin{cases} \displaystyle\bigoplus_{i=0}^{n} \bigoplus_{j=0}^{m-1} H_{i,[k+j]_m} & \text{if } \Psi(H) = 1 \\ \\ \displaystyle H_{0,k} \bigoplus_{i=1}^{n} \bigoplus_{j=0}^{m-1} H_{i,[k+j]_m} & \text{otherwise} \end{cases}$$

where $H_{0,k} = \bigoplus_{i=0}^{m-1} h_{0,d}^{[i+k+d]_m}$, as shown in Fig. 2. The distance between two feature H and $\overline{H}$ is then given by

$$\widehat{d}(H,\overline{H}) = \min_{k=0,\ldots,m-1} d(H,\overline{H}_{\alpha k})$$

where $d(\cdot,\cdot)$ is a common distance measure like L_1 or L_2 and each descriptor vector has been normalized to unit length.

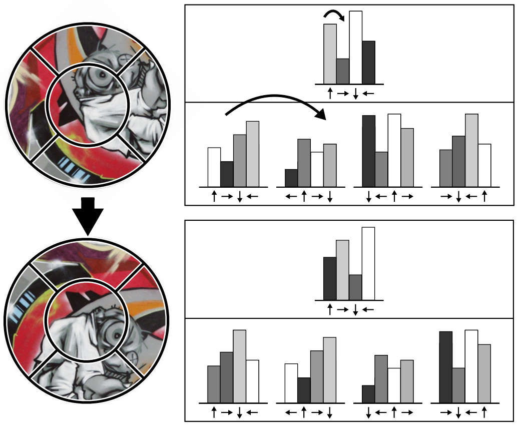

Figure 2. Discrete rotation on the descriptor for $\Psi(H) = 0$, $n = 1$ and $m = 4$.

We developed also a further variant of the sGLOH descriptor, called sGLOH+, which computes a small orientation refinement on the patch before computing the descriptor vector. Instead of estimate the dominant orientation on the whole range $[0, 2\pi[$ (see Fig. 3, left), the dominant orientation is computed using modular arithmetic on the range $[0, z[$, where $z = \frac{2\pi}{m}$ (see Fig 3, right). The histogram for the dominant orientation computation, if v bins are required, is then given by

$$t^i = \frac{1}{\sqrt{2\pi}\overline{\sigma}} \sum_{p\in R} G_m(p)\; e^{-\frac{(M_z(S_m(G_d(p))-\overline{m}_i))^2}{2\overline{\sigma}^2}}$$

where R is the feature patch, $i = 0,\ldots,v-1$, $\overline{\sigma} = \frac{z}{v}c$, $\overline{m}_i = \frac{z}{v}i$ and $S_m(x) = x - z\lfloor\frac{x}{z}\rfloor$ is the remainder, considering a period of length z. The dominant orientation t is then given by

$$t = \frac{z}{v} \operatorname*{argmax}_{i=0,\ldots,v-1} t^i$$

Since in this case the range of possible orientations is more costrained, the accuracy of the matching process increases. Moreover, in the case of bad dominant gradient orientation estimation, the error is bounded by $[0, \frac{z}{2}]$, using the distance measure $\widehat{d}$.

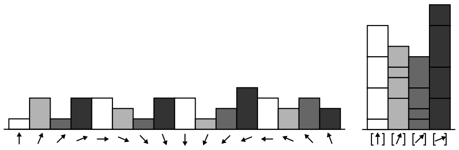

Figure 3. Dominant gradient orientation histogram computation for $m, v = 4$.

3. Results

The sGLOH and sGLOH+ descriptors have been compared with the SIFT descriptor on the well-known Oxford database [1]. The same experimental setup described in [13] has been used, while keypoints have been extracted with the Harris-Z [3] detector which was proved to provide stable features. The first and fourth images of the graffiti and wall sequences have been used to check performance on viewpoint changes, while the first and the fourth images of the bark and boat sequences have been adopted to test the stability of the descriptors on scale and rotation changes. The recall is defined as

$$recall = \frac{\#correct\ matches}{\#correspondences}$$

while the precision is given by

$$precision = \frac{\#correct\ matches}{\#total\ matches}$$

The number of correct matches and correspondences is computed according to the overlap error [14], defined as the ratio of the intersection and union of the regions. As in [13], the overlap error is fixed to $\varepsilon = 0.5$, the feature patch is 41×41 pixels, while the nearest neighbour matching strategy is adopted. The sGLOH and sGLOH+ descriptors have been tested for $c = 0.7$, $m = 8$, $n = 0, 1, 2$, so that the descriptors grid radii are respectively $\{20\}, \{12, 20\}, \{7, 13, 20\}$ pixels. When $n = 1, 2$, also the different cases $\Psi(H) = 0, 1$ have been examined, obtaining descriptors of lengths $l = 64, 72, 128, 136, 192$. For the sGLOH+ descriptor the dominant orientation is estimated inside a disk centered in the patch with 8 pixels radius as in [1] and the precision for the gradient estimation v has been set to 8.

The best distance measures between L_1, L_2 have been used for each descriptor. The sGLOH and sGLOH+ descriptors perform best on the L_1, while the L_2 distance was used in the case of the SIFT descriptor.

Plots are shown in Fig. 4. The obtained results are comparable with those obtained by SIFT (blue surface).

Only in the case of the bark sequence, SIFT performs better, but only for really high precision. It can be seen that both sGLOH and sGLOH+ detectors presents local minima for $l = 72, 136$, underlining better result for the configurations with $\Psi(H) = 1$. Moreover, sGLOH+ (green surface) seems to perform better than sGLOH (red surface) in most cases, especially when the precision is low, underlining the validity of the dominant orientation refinement introduced.

The sGLOH and sGLOH+ descriptors do not introduce any relevant computational costs in the generation of the descriptor vector. The new distance measure $\widehat{d}$ is more time consuming, but still acceptable compared to the L_2 distance when the L_1 distance, which performs better on the new descriptors, is used.

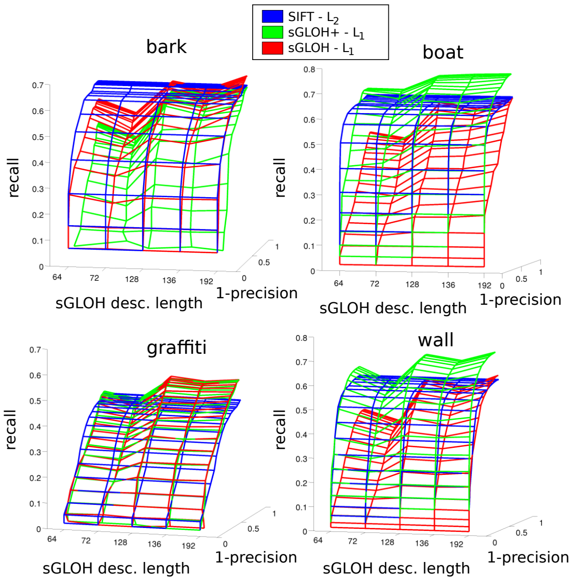

Figure 4. Recall/1-precision for test images.

4. Conclusion

In this paper we present two modification on the GLOH descriptor, called sGLOH and sGLOH+, which avoid to rotate the feature patch before computing the descriptor since predefined discrete orientations can be easily derived by shifting the descriptor vector.

Moreover, the method is refined in the sGLOH+ descriptor, by estimating tha dominant orientation by modular arithmetic in a defined value range, bounding the error.

The proposed descriptors have been compared with the SIFT descriptor on the Oxford image dataset, with good results in terms of robustness and stability, at a reasonable increase in the computational cost.

References

[1] Affine covariant features, 2007. http://www.robots.ox.ac.uk/∼vgg/research/affine.

[2] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[3] F. Bellavia, D. Tegolo, and C. Valenti. A non-parametric scale-based corner detector. In *International Conference on Pattern Recognition*, 2008.

[4] Y. Cui, N. Hasler, T. Thormählen, and H. Seidel. Scale invariant feature transform with irregular orientation histogram binning. In *International Conference on Image Analysis and Recognition*, pages 258–267, 2009.

[5] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[6] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition, 2004.

[7] Y. Ke and R. Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition*, pages 506–513, 2004.

[8] T. Lindeberg. Feature detection with automatic scale selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[9] H. Ling and K. Okada. Diffusion distance for histogram comparison. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 246–253, 2006.

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

[11] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *British Machine Vision Conference*, pages 384–393, 2002.

[12] K. Mikolajczyk and C. Schmid. Scale and affine invariant interest point detectors. *International Journal of Computer Vision*, 60(1):63–86, 2004.

[13] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005.

[14] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1/2):43–72, 2005.

[15] M. Toews and W. Wells III. Sift-rank: Ordinal description for invariant feature correspondence. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2009.

[16] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009.