Efficient Rotation-Discriminative Template Matching

David Marimon and Touradj Ebrahimi

Ecole Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{david.marimon,touradj.ebrahimi}@epfl.ch

Abstract. This paper presents an efficient approach to rotation discriminative template matching. A hierarchical search divided in three steps is proposed. First, gradient magnitude is compared to rapidly localise points with high probability of match. This result is refined, in a second step, using orientation gradient histograms. A novel rotation discriminative descriptor is applied to estimate the orientation of the template in the tested image. Finally, template matching is efficiently applied with the estimated orientation and only at points with high gradient magnitude and orientation histogram similarity. Experiments show a higher performance and efficiency as compared to similar techniques.

Keywords: template, matching, rotation, gradient, histogram.

1   Introduction

Visual matching consists in comparing the visual information extracted from an image patch with the same type of information extracted from another image. Applications that use visual matching are related to region recognition, for instance, object tracking [1, 2, 3].

Existing techniques can be divided into two categories: trained and non-trained. For the first group, classifiers are trained with a test set of positive and negative patch examples. Research is concentrated on the data set and the classification techniques. These techniques provide an excellent compromise between speed and performance at run-time [4]. However, the time consumed to gather or generate the training data and to train the classifier is generally high. For the second group, attention is paid on the description of the information rather than in the training data or the classification scheme used. Most researchers have focused on descriptors and measures that are robust or even invariant to viewpoint and/or illumination distortions [5]. In this case, descriptors are generally built from a single instance of the patch to recognise. The drawback of such invariance is often a higher computational cost during the matching process.

In this paper, we concentrate on this second category and target fast computation environments such as tracking applications. For this goal, we propose a rotation-discriminative patch descriptor and an efficient hierarchical search strategy divided in three steps. First, similar gradient magnitude is exhaustively

searched within the image. The most similar points are sorted out. Second, the orientation gradient histogram is matched at those points providing a measure of similarity, together with an estimate of the rotation that the patch has undergone. Again, only the most similar points are kept. Finally, template matching is performed at those points by computing the Normalised Cross Correlation (NCC) between the intensity neighbourhood of the point in the image and the patch rotated according to the orientation estimated in the previous step.

2   Related Works

This section takes a more detailed look into non-trained matching techniques related to the proposed method. The matching process is done by comparing the descriptor of a patch with the descriptors obtained at different locations in an image. This description determines in general the robustness of a recognition process facing viewpoint and illumination changes. Consequently, most researchers concentrate their efforts on obtaining invariant descriptors. The reader can find a comparison of descriptors in [5]. Among those descriptors and the related recognition strategies, some have been chosen according to their relation to the method proposed here, for more in depth explanation.

Two descriptors have been used extensively for recognition purposes: templates and distributions. *Templates* are ordered arrays of the pixel values of an image region and have two main advantages. First, the simplicity of construction of this descriptor. Second, that the spatial information of the region is maintained. The drawback is the high sensitivity to viewpoint and illumination changes. Several improvements of template matching techniques exist in literature, either concentrating in illumination changes [1] or also in geometrical variations [2]. *Histograms*, are arrays that model the true distribution by counting the occurrences of pixel values that fall into each bin (range of values). Different information can be used for histogram descriptors, e.g. gray-scale, colour [6,3], and gradient [7]. Histograms have opposite advantages and drawbacks when compared to templates. In other words, histograms loose spatial information while viewpoint invariance can be achieved by construction. Several attempts at combining spatial and distribution information exist, e.g., [8,6,3,7]. Among them, we emphasise a convex monotonic decreasing kernel [3] that weights the contribution of pixels to the histogram. This kernel lessens the weight of peripheral pixels which are the least reliable, being often affected by occlusion, background and viewpoint changes. Also relevant is the use of spatial distribution of gradient histograms achieving high viewpoint invariance [7].

The strategy to locate and match regions inside an image varies depending on the application and often also on the complexity of the descriptor. In applications such as *point correspondence* [7], only locations with high repeatability are considered. Once the detection of possible candidates (usually a large amount of points) in each image is performed, a pair-wise match has to be set. In *line-search* matching, the goal is to iteratively maximise the similarity between the patch and different points of an image. At each iteration, a new position sensed

to increase the similarity is found using, for instance, gradient information [3]. In *window-search* (or exhaustive) matching, the similarity is computed at each point in a test image. As the computational power needed is proportional to the size of the image, it is often applied only when the descriptor is computed rapidly or the size of the image is relatively small [1].

Examples of fast rotation invariant template matching with an exhaustive search are [6, 9]. Fredriksson et al. [6] use an orientation invariant descriptor (colour histogram), to locate points with high probability of match. Although this method is faster than cross correlation by FFT, histograms are not efficiently computed in this work. Ullah et al. [9] presented a two step strategy. First, orientation code histograms (OH) are used to estimate the orientation of a patch in each point of an image. Second, orientation code matching (OCM) at the right orientation is applied only to the best histogram matches. This independent work differentiates from the method proposed here in two main contributions. First and most important, the OC is built only upon the extracted patch at a single orientation achieving less invariance to rotations than our descriptor. Second, that the processing time needed to produce a match is much higher (see Sect. 4).

3   Proposed Method

The problem that we are tackling is template matching of patches that have undergone rotations. A straight approach to this problem would be to generate a number of rotated versions of a patch and to correlate them at each point of the tested image. This window matching process has however a high computational cost. Instead, we propose to estimate first which rotated version has the highest probability of being the adequate to maximise the level of correlation. This is done by comparing the orientation gradient histogram of the patch and that of the neighbourhood extracted at several points in the image. In order to perform most scan operations rapidly we take advantage of the integral image (running sum of image) [4] and the integral histogram (running sum of bins) [10]. In this section, the descriptor used for the recognition of a patch and the proposed matching are detailed.

3.1   Region Description

Gradient information is chosen to generate the descriptor of a patch. One of the reasons lies on the little sensitivity of the gradient to illumination changes, which is one of the problems that recognition has to deal with. As described in Sect. 1, another major problem to tackle is viewpoint invariance. We propose a descriptor that deals with this problem concentrating on rotation robustness and, at the same time, provides orientation information of the region it describes.

Let us first analyse the behaviour of the gradient. From a theoretical point of view, the gradient has a continuous response to a continuous and derivable function. Suppose that a gradient orientation histogram of N bins is computed from a patch $\mathbf{P}$. In this case, a rotation of the patch by δ degrees changes the

values in the histogram. In particular, when $\delta = n \cdot 360/N$ where $n \in \mathbb{Z}$, the histogram would be exactly equal to a perfect shift, and the shift in bins would be equal to n. However, this ideal case is not fulfilled in reality.

Following the observation that histograms change with different orientations, we propose to generate rotated versions of a patch and, from these versions, create a single histogram that can deal with rotations. As mentioned before, orientation histograms repeat approximately their shape every $\Delta = 360/N$ degrees. This can be exploited by aligning the histograms of versions rotated exactly by $k\Delta$ with $k \in \mathbb{Z}$.

The histogram descriptor is obtained as explained next. Firstly, N rotated versions of the patch $\mathbf{P}$ to be matched are pre-computed with an angle of rotation of $n\Delta$ degrees (for $n = 0, .., N-1$) where N is the number of bins. These versions are cropped so as to eliminate additional pixels introduced by the rotation, leading to a vector of rotated versions of the patch $\overrightarrow{\mathbf{P}}_i$, where i indexes the vector. Secondly, the gradient of each of these versions is computed at each point (x, y) as follows

$$
\begin{aligned}
dy(x,y) &= \overrightarrow{\mathbf{P}}_i(x, y+1) - \overrightarrow{\mathbf{P}}_i(x, y-1) \\
dx(x,y) &= \overrightarrow{\mathbf{P}}_i(x+1, y) - \overrightarrow{\mathbf{P}}_i(x-1, y) \\
\nabla_m(x,y) &= \sqrt{dy(x,y)^2 + dx(x,y)^2} \\
\nabla_\theta(x,y) &= \arctan(dy(x,y), dx(x,y)),
\end{aligned}
\tag{1}
$$

where $\arctan(a,b)$ is a function that returns the inverse tangent of $\frac{a}{b}$ in a range $[0, 2\pi]$, ∇_m is the magnitude and ∇_θ is the orientation of the gradient. Then, ∇_θ is quantised in N bins. In order to compact the statistical description of the patch and to reduce the effect of noise, the contribution of each point in $\nabla_\theta(x, y)$ to the corresponding bin is weighted by its magnitude $\nabla_m(x, y)$ (similar to the approach in [7]). It is desirable that the weight of the peripheral pixels is lessened. However, applying a kernel (as presented in [3]) is not possible with the integral histogram approach. We approximate the effect of the kernel by giving double weight to the central part of the patch. Finally, the global histogram of the patch is the mean obtained with the N histograms aligned according to their rotation. Fig. 1 shows an example for 16 bins with the original patch and its rotated versions with the corresponding histogram aligned accordingly.

This average of rotated versions gives a robust descriptor when the rotation of the image is around $n\Delta$ degrees. It could be argued that for non-integer bin-wide angles higher variations will occur. However, experiments shows that, with enough bins, this descriptor is reliable even around $n\Delta + \Delta/2$ degrees (see Sect. 4).

The final region descriptor is composed of the global histogram $\tilde{\mathbf{h}}$, its variance σ^2, its norm and the rotated versions of the template. Our experimentation has shown that using the variance in the matching process enhances the performance.

$$
f(\mathbf{P}) = \left[\tilde{\mathbf{h}}_\mathbf{P}, \sigma^2_\mathbf{P}, \|\tilde{\mathbf{h}}_\mathbf{P}\|, \overrightarrow{\mathbf{P}}_0, \dots, \overrightarrow{\mathbf{P}}_{N-1} \right].
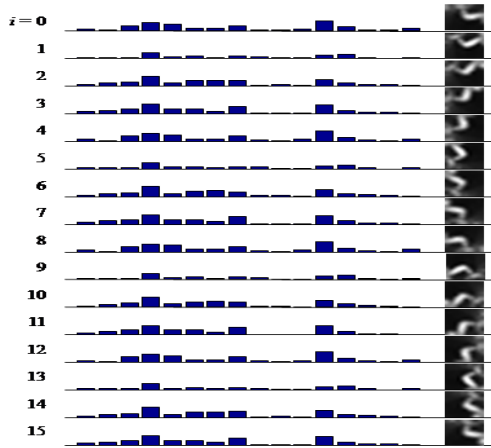\tag{2}
$$

Fig. 1. Example of histogram alignment with $N = 16$ bins. Central column: histograms aligned according to their rotation; right column: corresponding original patch ($i = 0$) and rotated versions ($i = 1, \ldots, 15$).

3.2 Rotation-Discriminative Template Matching

This section describes the three hierarchical selection steps performed. Firstly, an exhaustive gradient magnitude comparison is performed. Secondly, the candidates with highest magnitude similarity are kept for orientation gradient histogram matching. This matching provides also an estimate of the rotation between the patch and the image. Finally, template matching is performed at the position of the most similar histograms using the rotation estimated previously.

Gradient magnitude matching. The norm of the histogram $\|\tilde{\mathbf{h}}_{\mathbf{P}}\|$ can be used as a simple feature to rapidly scan the image for similar candidates. From the construction of the histogram it can be found that $\|\tilde{\mathbf{h}}_{\mathbf{P}}\| \simeq \sum_{\mathbf{P}} \nabla_m + \sum_{\mathbf{P}'} \nabla_m$, where $\mathbf{P}'$ is the central part of the patch. Following this observation, we propose to compare this norm with each neighbourhood in a window-search strategy. This can be efficiently performed with the integral image [4] of the magnitude gradient. Given a neighbourhood $\mathbf{R}$ of a point, the measure used to compare the norm is $d_m = \exp{-\alpha \left(1 - \left(\sum_{\mathbf{R}} \nabla_m + \sum_{\mathbf{R}'} \nabla_m\right) / \|\tilde{\mathbf{h}}_{\mathbf{P}}\|\right)^2}$, where α is a factor that weights this similarity according to the variance of the histogram. More precisely, $\alpha = N/(1000 \cdot \|\sigma_{\mathbf{P}}^2\|)$. The points in the image that have a similarity $d_m > 0.9$ are kept as candidates for further matching.

   In the worst case where similar magnitude is found all over the image, the number of candidates remains the same after this step. However, based on experiments, this simple selection criteria permits a reduction of the number of candidates by an average factor of 20.

Histogram matching. The gradient orientation histogram matching is applied to the candidates with similar histogram norm. Histogram can be efficiently

computed with the integral histogram approach [10]. The gradient orientation histogram of a region in the image is obtained from the contribution of the quantised $\nabla_\theta(x, y)$ weighted with $\nabla_m(x, y)$ (as for the descriptor of the patch).

The similarity between the histogram of the patch $\tilde{\mathbf{h}}_{\mathbf{P}}$ and that of each candidate is computed with a custom measure to compare orientation histograms (or, generically, circular vectors), the *Circular Normalised Euclidean Distance* (CNED). Not only the CNED measures the distance d between two vectors, but it also determines the circular shift $\hat{s}$ that corresponds to the minimal distance. Mathematically expressed $\mathbf{CNED}(\mathbf{a}, \sigma_{\mathbf{a}}^2, \mathbf{b}) = [\hat{s}(\mathbf{a}, \sigma_{\mathbf{a}}^2, \mathbf{b}) \quad d_{\hat{s}}(\mathbf{a}, \sigma_{\mathbf{a}}^2, \mathbf{b})]^T$, where $\hat{s}(\mathbf{a}, \sigma_{\mathbf{a}}^2, \mathbf{b}) = \arg\min_s d_s(\mathbf{a}, \sigma_{\mathbf{a}}^2, \mathbf{b})$ and

$$d_s(\mathbf{a}, \sigma_{\mathbf{a}}^2, \mathbf{b}) = \sqrt{\sum_{i=0}^{N-1} \frac{(\mathbf{a}(i) - \mathbf{b}((i+s) \bmod N))^2}{\sigma_{\mathbf{a}}^2(i)}}, \tag{3}$$

where $\mathbf{a}$ and $\mathbf{b}$ are vectors of length N, s is the shift that takes a discrete value between 0 and $N-1$, mod is the modulus function, and $\sigma_{\mathbf{a}}^2$ is the variance associated to vector $\mathbf{a}$. The result of this matching is hence a similarity score $d_{\hat{s}}$ and an estimate of the orientation of the patch $\hat{s} \cdot \Delta$ for each candidate.

Template matching. The magnitude and the orientation histogram discard many unrelated points but the result is still not selective enough (as seen below in Sect. 4). The template is used as a further selection criterion. More precisely, template matching is done using a Normalised Cross Correlation (NCC) between the templates $\mathbf{R}$ centered at those points with high histogram similarity and the template of the patch $\vec{\mathbf{P}}_{\hat{s}}$.

$$NCC\left(\vec{\mathbf{P}}_{\hat{s}}, \mathbf{R}\right) = \frac{\sum\sum(\mathbf{R} - \overline{\mathbf{R}}) \cdot (\vec{\mathbf{P}}_{\hat{s}} - \overline{\vec{\mathbf{P}}_{\hat{s}}})}{\sqrt{\sum\sum(\mathbf{R} - \overline{\mathbf{R}})^2 \cdot \sum\sum(\vec{\mathbf{P}}_{\hat{s}} - \overline{\vec{\mathbf{P}}_{\hat{s}}})^2}}, \tag{4}$$

where $\overline{R}$ is the average value of R. By subtracting this mean value, the result is invariant to uniform illumination changes.

4   Experiments

This section assesses the performance of the proposed method in comparison to other similar techniques. Firstly, the techniques compared are described. Secondly, the set of test images and image patches are presented. Thirdly, the evaluation methodology is explained. Finally, results are depicted and discussed.

The matching techniques compared are: the NCC computed for all the N rotated versions at each point (we call this NCC-R), a gray-level intensity histogram matching (IHM), our own gradient orientation histogram matching (GHM) but computed exhaustively in the image (this is, without magnitude pre-sorting), the technique presented in [9] (OH+OCM), and the final correlation result of our method. In order to do a fair comparison with the IHM, a

Fig. 2. Original images used for the experiments. Average size: 300x225 pixels.

N-bin histogram of the intensity values of each rotated version is computed for each patch. This gives N histograms which are averaged bin-by-bin into a single intensity histogram describing the patch. Moreover, the central part of each histogram is given more weight as for $\mathbf{h_P}$ (see Sect. 3.1). The similarity measure used in this case is the Euclidean distance.

The set of images used for testing is shown in Fig. 2. The first two images (top-left) are custom whereas the other six images are taken from the Visual Geometry Group database [11].

There is one key parameter in the method: the number of bins N in the histogram. This number determines the value of $\Delta = 360/N$ and hence the performance of the method. More concretely, the whole matching is expected to work better for rotations around $k\Delta$ than around $k\Delta+\Delta/2$ (with $k = 0, .., N-1$). Experiments are run on 10, 16 and 20 bins to give an approximate idea of a lower and upper performance bounds. The images are rotated 20 and 70 degrees for a histogram of 10 bins ($\Delta = 36^o$), and 10 and 70 degrees for both 16 bins ($\Delta = 22.5^o$), and 20 bins ($\Delta = 18^o$). These angles correspond to almost best and worst case scenarios for each histogram length.

For each one of the original images, a set of patches is extracted. Their sizes range from 10x10 to 20x20 pixels, which is a common range in related research. Around 20 patches per image are extracted with the Harris corner detector [12]. The main reason behind the choice of this point detector is that patches have more relevant texture information in this case.

The purpose of the compared methods is to find matches. A correct match is found when a point with high similarity coincides with the ground truth. This idea is translated into the concept of *true positive* and of *false positive* in the opposite case. The performance of matching technique can be given by these two values. More precisely, the higher the number of true positives and lower the number of false positives, the better is the result.

The level of similarity that determines a match (or positive) is given by a range, i.e. maximum to minimum similarity, which is not the same for all the considered techniques. Nevertheless, it is possible to find a range that varies

equivalently. In order to find this equivalence, the values of each similarity map are taken, independently, in descending order (highest to lowest similarity) regardless of the value itself. An equivalent level of similarity is found in this case by parsing each list of values. The true positives and false positives can then be defined mathematically. Assume that $d(f_1(a), f_2(b))$ is the similarity between two descriptors $f_1(a)$ and $f_2(b)$ of the respective regions a and b. Given an image $\mathbf{I}$, a rotated version $\hat{\mathbf{I}}$, a patch $\mathbf{P}$ extracted from $\mathbf{I}$ at $(x_\mathbf{P}, y_\mathbf{P})$, and $(\hat{x}_\mathbf{P}, \hat{y}_\mathbf{P})$ being the correspondence of $(x_\mathbf{P}, y_\mathbf{P})$ into $\hat{\mathbf{I}}$, a *true positive* is

$$\text{tp}_{\mathbf{P},\hat{\mathbf{I}},t} = \begin{cases} 1 \text{ if } \exists (x, y) \in \mathbf{G} \mid d(f_1(\mathbf{P}), f_2(\hat{\mathbf{R}}_{x,y})) > t \\ 0 \text{ otherwise,} \end{cases} \tag{5}$$

and conversely, a *false positive* is

$$\text{fp}_{\mathbf{P},\hat{\mathbf{I}},t}(x, y) = \begin{cases} 1 \text{ if} (x, y) \notin \mathbf{G} \mid d(f_1(\mathbf{P}), f_2(\hat{\mathbf{R}}_{x,y})) > t \\ 0 \text{ otherwise,} \end{cases} \tag{6}$$

where $t \in [\max d(f_1(\mathbf{P}), f_2(\hat{\mathbf{R}}_{x,y})), \min d(f_1(\mathbf{P}), f_2(\hat{\mathbf{R}}_{x,y}))]$ and $\mathbf{G}$ is the region $\{\hat{\mathbf{I}}(x, y) | x = \hat{x}_\mathbf{P} \pm 1 \text{ and } y = \hat{y}_\mathbf{P} \pm 1\}$. A 1 pixel neighbourhood is set to account for sub-pixel location after the image transformation.

The response of each matching method for the best 500 matches is depicted in Fig. 3. The NCC-R indicates a great performance almost independent of the number of candidates. This shows the high selectivity of this kind of map. In the case of the IHM, rotation invariance is evidenced by very similar results throughout the different cases. A poorer selectivity is shown by the GHM as a large number of false positives is obtained in order to get a high probability of having a true positive. The OH+OCM [9] has lower performance probably due to its non-invariant nor robust descriptor. Using only a single version to build the histogram is not enough to effectively face the variations in the histogram due to rotations. The results of the GHM are greatly improved when used as an input for the further template matching step of our method (especially visible as the number of bins grows). Furthermore, the proposed descriptor and similarity measure achieve the desired rotation discrimination and hence accurate matching.

Computational complexity. The efficiency of these methods is contrasted here with the processing time needed to produce a match. Table 1 shows this time (averaged for the patches in the test set) when computed with a Pentium M Processor at 1700 MHz. As it can be seen, the slowest algorithm is the OH+OCM. The main reasons are the circular mask used for matching and, consequently, the impossibility of using the integral histogram approach. The NCC uses the integral image as in [1]. Despite this fast matching implementation, it can be seen that comparing each rotated version of the template is inefficient. The estimation of the orientation in the histogram matching step drastically palliates this inefficiency. Moreover, the hierarchical selection proposed in our work enables a processing time almost as fast as the most simple and efficient
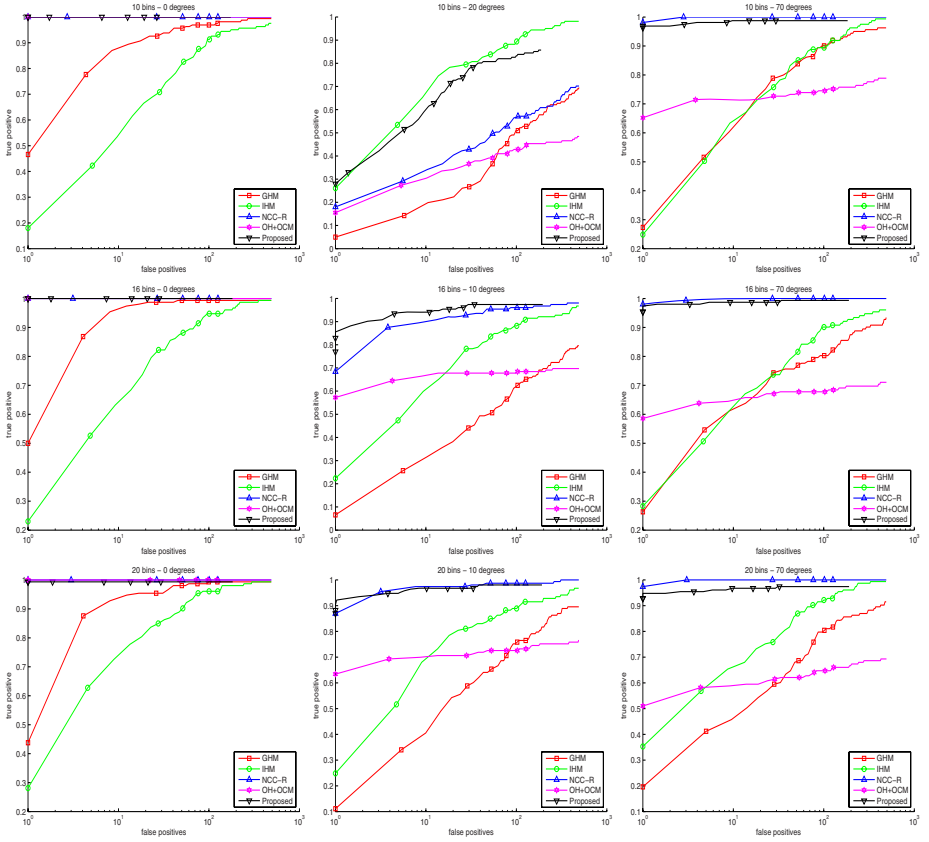
Fig. 3. Mean true positive (vertical axis) and false positives (horizontal) among all the patches. Number of bins: 10 (top), 16 (middle) and 20 (bottom). Rotation angle: 0 degrees (left), $\sim k\Delta + \Delta/2$ (central) and $\sim k\Delta$ (right).

Table 1. Average processing time for a single patch

	10 bins [s]	16 bins [s]	20 bins [s]
OH+OCM	4.4776	4.7115	4.9312
NCC-R	0.9001	1.2506	1.5235
GHM	0.4311	0.8816	1.3014
Proposed method	0.1773	0.2028	0.2584
IHM	0.1085	0.1275	0.1491

strategy, which is the IHM (implemented with the integral histogram). It should be pointed out that reducing the number of candidates would further reduce the computational cost.

5   Conclusions

An efficient method to perform rotation discriminative template matching has been presented. This is achieved with an orientation histogram matching followed by template matching. The main contributions of the method are a rotation-discriminative descriptor and the efficient matching strategy. Experimentation shows that our results are as good as performing NCC of each rotated version of a template but with an average speed up factor of six. In addition, performance and efficiency of our technique is superior to the most similar technique, namely, the OH+OCM [9]. Future path of research will focus on analysing the dependance of the method on the texture information available in the patch.

Acknowledgements

References

1. Lewis, J.: Fast template matching. Vision Interface, 120–123 (1995)
2. Hager, G., Belhumeur, P.: Real-time tracking of image regions with changes in geometry and illumination. In: Computer Vision and Pattern Recognition (CVPR), pp. 403–410 (1996)
3. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Trans. on Pattern Analysis and Machine Intelligence 25(5), 564–577 (2003)
4. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Computer Vision and Pattern Recognition (CVPR). vol. 1, pp. 511–518 (2001)
5. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)
6. Fredriksson, K., Ukkonen, E.: Faster template matching without FFT. In: Intl. Conf. on Image Processing (ICIP). vol. 1, pp. 678–681 (2001)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. Intl. Journal of Computer Vision 60(2), 91–110 (2004)
8. Haralick, R., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. on Systems, Man and Cybernetics 3(6), 610–621 (1973)
9. Ullah, F., Kaneko, S.: Using orientation codes for rotation-invariant template matching. Pattern Recognition 37(2), 201–209 (2004)
10. Porikli, F.: Integral histogram: a fast way to extract histograms in cartesian space. In: Computer Vision and Pattern Recognition (CVPR). vol. 1, pp. 829–836 (2005)
11. Visual Geometry Group, Robotics Research Group, Department of Engineering Science, University of Oxford, http://www.robots.ox.ac.uk/~vgg/data.html
12. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conf. pp. 147–151 (1988)