

ROTATION INVARIANCE IN IMAGES

Sam Mavandadi, Student Member, IEEE and Parham Aarabi, Senior Member, IEEE

Department of Electrical and Computer Engineering, University of Toronto

ABSTRACT

Rotation is one of the most basic transformations that can relate two images. To determine if two images are rotated versions of each other, one can either exhaustively rotate them in order to find out if they match up at some angle, or alternatively extract features from the images that can then be compared to make the same decision. In this paper, we will propose a novel method for extracting the components of an image that are invariant to rotation based on the Fourier Transform. We will compare the performance of the algorithm to the exhaustive search method and show that this is a much faster technique, that is also accurate in matching rotated images.

Index Terms— Image matching, Image Analysis, Image Processing, Image Texture Analysis, Image Retrieval, Rotation Invariance

1. INTRODUCTION

Invariance is a concept that is of extreme importance in many Image Processing applications. This importance has its roots in the way that we humans process visual information and analyze the images that are captured by our eyes. Invariance with respect to orientation, scale, and translation are essential for a human vision system to be able to function properly, recognizing objects, locations, and basically understanding images. In this paper, we will address the issue of invariance with respect to rotation from a signal processing perspective.

The issue of rotation-invariance has been extensively explored in the literature in the past. There has been work done in the optics community, the signal processing community [1, 2, 3], and the computer vision community [4, 5, 6, 7, 8, 9], and due to the different applications at hand, they have developed different methods for this purpose.

Probably the best-known signal processing approach to achieving rotation-invariance is the Fourier-Mellin Transform (FMT) [10, 11, 1]. This transformation involves taking the magnitude of the two dimensional Fourier Transform (FT) of the image, followed by a conversion to log-polar coordinates. This is then followed by a Mellin Transform (MT) which achieves invariance to rotation, translation, and scale. The applications in the computer vision community are more concerned with object recognition, and multi-view automatic

scene matching. As a result, more work is done toward developing *local* image features that are invariant to rotation, scale, and other transformations. The best-known work in this field is described in [9]. In this paper, Lowe develops a scale and rotation invariant local feature detector and also a descriptor that is invariant to both of these transformations. To achieve rotation invariance in the descriptor, a dominant orientation is determined by looking at the local image gradients at the point of interest, and the features in the descriptor are essentially rotated to offset this dominant orientation, resulting in a rotation-invariant descriptor. This method of using a dominant orientation for achieving invariance to rotation has become the method of choice by many [9, 7, 4].

There has also been some work in the area of affine-invariance, which deals with almost all types of image transformations that are possible under everyday conditions, and not just rotation and scale changes [4].

2. ROTATION INVARIANCE

2.1. Representation

The Cartesian coordinate system is very awkward for dealing with rotation. The polar coordinate system is a much more natural domain for such tasks. Once in the polar coordinate system, the rotation operation is transformed into a circular shift in the angle axis, and becomes much more easier to deal with. We thus describe the images, or image patches using a polar coordinate frame such that $\mathbf{p}(x, y) \rightarrow \mathbf{p}(r, \theta)$. For image patch \mathbf{p} , where $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$, $\theta = \arctan\left(\frac{y - y_0}{x - x_0}\right)$, and (x_0, y_0) is the center of the image patch.

2.2. Invariance

Every image patch contains components that are inherently invariant under rotation. The most obvious and simple of these components is the DC component of the pixels intensities. In other words, the average intensity. This intuitively makes sense and is obviously seen if we look at the Fourier representation of the patch. Another well-known rotation invariant representation of an image patch is its color-histogram. Such a histogram represents the distribution of different colors in the patch, and so would not be affected by any transformation that would preserve this distribution. In the absence

of noise, rotation would completely preserve the distribution of colors in an image patch, and thus its color-histogram.

The mentioned methods for patch representation however, allow for too many degrees of freedom. In other words, there are too many transformations and deformations of a patch that would yield the same feature, and this leads to the false matching of patches that are not the same. We will now describe a method for an image descriptor that attempts to only discard the rotation-variant components of the image.

2.2.1. Fourier representation of image patches

An equivalent method of representing the polar image patch would be to describe it using its Fourier coefficients at each radius. In other words, we can treat $\mathbf{p}(r, \theta)$ as separate signals for all values of r and take their Fourier Transforms along the θ axis. We then have

$$\mathcal{P}(r, \omega) = \int_{-\infty}^{\infty} \mathbf{p}(r, \theta) e^{-j\omega\theta} d\theta \quad (1)$$

Now, in the discrete case, this becomes

$$\mathcal{P}[n, k] = \sum_{l=0}^{N_\theta-1} \mathbf{p}_n[l] e^{-j2\pi l \frac{k}{N_\theta}} \quad (2)$$

where $\mathbf{p}_n[l]$ represents the patch at the n^{th} radial distance for discrete angle values l .

2.2.2. Rotation-Invariance through Shift-Invariance

As previously mentioned, with the conversion to polar representation, any rotation in the image patch is transformed into a circular shift along the θ axis. Thus, the task of determining the rotational transformation between two image patches is now a phase-difference-estimation problem which has been addressed for many other applications in array signal processing. Looking at the representation that we have for the image patch, we see that the phase of the Fourier representation contains the rotation information in the image, thus rotation would have no effect on the magnitude. We have

$$|\mathcal{F}(\mathbf{p}_r(\theta - \theta_0))| = |e^{j\omega\theta_0} \mathcal{P}_r(\omega)| = |\mathcal{F}(\mathbf{p}_r(\theta))| \quad (3)$$

So, in the continuous case, any linear shift along the θ direction (i.e. rotation) would be discarded. In the discrete case, this linear shift becomes a circular shift:

$$\begin{aligned} |\mathcal{F}(\mathbf{p}_n[l - m])| &= \left| \sum_{k=m}^{m+N_\theta-1} \mathbf{p}_n[k - m] e^{-j2\pi \frac{k}{N_\theta}} \right| \quad (4) \\ &= |e^{-j2\pi \frac{m}{N_\theta}} \mathcal{F}(\mathbf{p}_n[l])| = |\mathcal{F}(\mathbf{p}_n[l])| \quad (5) \end{aligned}$$

We now have a descriptor that is invariant to rotation. This descriptor however has one major problem. The DC components of the Fourier magnitudes are much more pronounced than those of the higher frequency bins.

2.3. Frequency Scaling

As mentioned before, the current descriptor has a large bias at DC. This bias makes it difficult to distinguish between unrelated patches. In order to emphasize the effects of the higher frequency components, we will scale the frequencies such that the higher frequencies are weighted more than the lower frequency components. In our experiments, we use a logarithmic scale, which was empirically determined to give good results.

3. PHASE ENTANGLEMENT

3.1. Problem Statement

There is still one major problem with the derived rotation-invariant descriptor. Too much information is lost when we ignore the phase. The phase in fact contains more information than just the rotation angle between the two compared patches, which is the overall phase difference.

When phase is ignored as described in the previous section, too much rotational freedom is allowed in the patches. Basically, radial components (circularly sampled signal at each radius) are allowed to have independent rotations. For example, the two patches in Figure 1 would yield the same descriptor under the current formulation. This allows so much freedom, that even unrelated patches can yield high correlation values. Mathematically, we have the following:

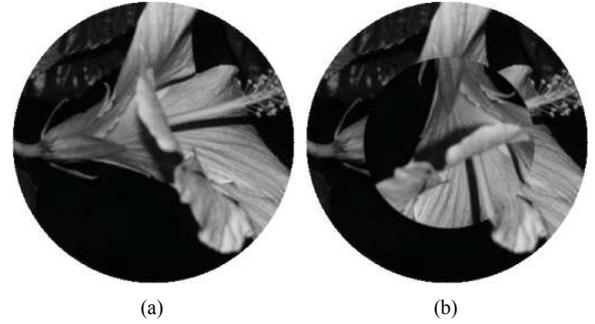


Fig. 1. The two image patches yield identical descriptors when phase is completely ignored.

$$\mathbf{p}[n, l] = \begin{pmatrix} \mathbf{p}_1[l] \\ \vdots \\ \mathbf{p}_{N_R}[l] \end{pmatrix} \quad (6)$$

where $\mathbf{p}[n, l]$ is the matrix containing the polar representation of the image patch and $\mathbf{p}_i[l]$ represents the sample of this patch at radius i . N_R is the number of discrete values at which the patch radius is measured. Then let $\mathcal{P}[n, k]$ be the matrix containing the Fourier coefficients of $\mathbf{p}[n, l]$. Let us now denote the matrix containing the magnitudes of these Fourier

coefficients as $|\mathcal{P}[n, k]|_{N_R \times \Omega}$, where Ω is the number of discrete frequency bins. This constitutes our rotation-invariant descriptor. Now, imagine that we have two patches \mathbf{p} and \mathbf{q} , such there are independent rotations ϕ_i at every radial value i between the two images. So,

$$\mathbf{q}[n, l] = \mathbf{p}[n, l - \phi[n]] \quad (7)$$

where $\phi[n]$ is the discrete radius-dependent rotation (in samples). And

$$|\mathcal{Q}[n, k]| = \begin{pmatrix} |e^{-j2\pi\frac{\phi_1}{\Omega}} \mathbf{P}_1[l]| \\ \vdots \\ |e^{-j2\pi\frac{\phi_{N_R}}{\Omega}} \mathbf{P}_{N_R}[l]| \end{pmatrix} = |\mathcal{P}[n, k]| \quad (8)$$

And thus the descriptors are equivalent.

3.2. Entanglement

We would like to remove all the rotational information between the two images, and yet preserve the phase relationship that exists at different radial values. In other words, we want the descriptor to change significantly if we have a rotation at some radii and not the whole patch. This implies that we should create a descriptor where independent phase changes would also affect the magnitudes, unless the phase change is uniform across all radii.

To achieve this, we note that if we add two signals, then a linear shift in either one would change the magnitude of the Fourier representation of the summation. Similarly, if we add the different radial samples of a patch, then a shift in one would affect the magnitude of the descriptor. So, instead of using the simple polar representation of the image patch to compute the rotation-invariant descriptor, we will use linear combinations of the rows of the polar representation instead. We thus have $\psi_i[l] = \sum_{j=0}^{N_R} W_{ij} \mathbf{p}_j[l]$, where $\psi_i[l]$ is a linear combination of the rows \mathbf{p}_j of the original image patch $\mathbf{p}[n, l]$. This can be formulated very conveniently in matrix notation as $\boldsymbol{\psi} = \mathbf{W}\mathbf{p}$, where \mathbf{W} is a weight matrix of size $N_R \times N_R$. Note that in order to preserve the most amount of information, \mathbf{W} must have full rank to produce N_R linearly independent row vectors. Now if we have $\boldsymbol{\Psi} = \mathcal{F}\{\mathbf{W}\mathbf{p}\}$ and $\boldsymbol{\Gamma} = \mathcal{F}\{\mathbf{W}\mathbf{q}\}$, then

$$|\boldsymbol{\Gamma}_n[\mathbf{k}]| = \left| \sum_{l=1}^{N_R} W_{nl} \mathbf{Q}_l[k] \right| \quad (9)$$

$$|\boldsymbol{\Psi}_n[k]| = \left| \sum_{l=1}^{N_R} W_{nl} e^{-j2\pi\frac{\phi_l}{\Omega}} \mathbf{Q}_l[k] \right| \quad (10)$$

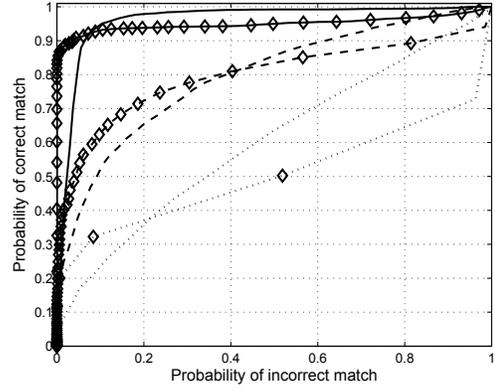


Fig. 2. Performance of the proposed algorithm vs. exhaustive search. The curves with the diamonds are those of the exhaustive technique, and the ones without, are those of the proposed method. The solid lines correspond to an added Zero-Mean Gaussian noise with variance of 0.001, the dashed lines correspond to that of variance 0.01, and the dotted line corresponds to a variance of 1.

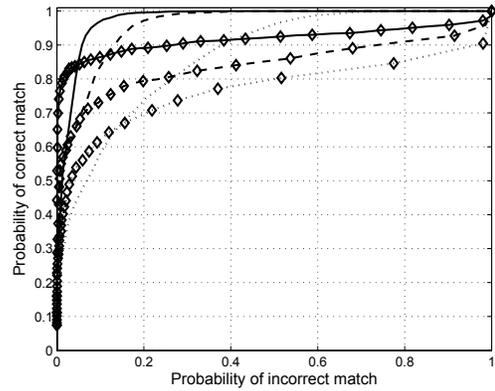


Fig. 3. Performance of the proposed algorithm vs. exhaustive search. The curves with the diamonds are those of the exhaustive technique, and the ones without, are those of the proposed method. The solid lines correspond to an added Speckle noise with variance of 0.016, the dashed lines correspond to that of variance 0.128, and the dotted line corresponds to a variance of 0.512.

Now if $\phi_1 = \phi_2 = \dots = \phi_{N_R} = \phi$, then we have

$$|\boldsymbol{\Psi}_n[k]| = \left| \sum_{l=1}^{N_R} W_{nl} e^{-j2\pi\frac{\phi}{\Omega}} \mathbf{Q}_n[k] \right| \quad (11)$$

$$= \left| e^{-j2\pi\frac{\phi}{\Omega}} \sum_{l=1}^{N_R} W_{nl} \mathbf{Q}_n[k] \right| = |\mathbf{Q}_n[k]| \quad (12)$$

We now have an entangled set of phases for each radius, which ensures that the descriptor remains the same only if there is an overall rotation of the image patch.

4. EXPERIMENTAL RESULTS

To test the performance of the proposed rotation-invariant transformation, the algorithm was run on a database of 512 im-

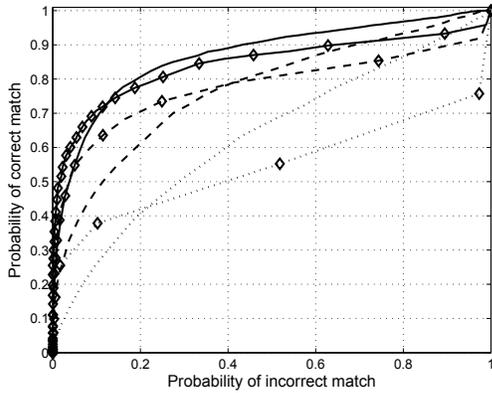


Fig. 4. Performance of the proposed algorithm vs. exhaustive search. The curves with the diamonds are those of the exhaustive technique, and the ones without, are those of the proposed method. The solid lines correspond to an added Salt & Pepper noise with variance of 0.1, the dashed lines correspond to that of variance 0.3, and the dotted line corresponds to a variance of 0.6.

ages of size 100×100 and 9 different rotation angles at increments of 40 degrees. The correlation coefficients of the comparisons of the different rotation angles (of the same image), and their correlation with other random images (taken from the database) were used as a measure of the algorithm's performance. This is then compared to the performance of an exhaustive search algorithm, where the two images are exhaustively rotated to achieve the highest possible correlation value. The correlation coefficient threshold is varied in order to generate the ROC curves for the full range of match and mismatch probabilities. Note that in a practical environment, it may not be possible to rotate the two images at all angles, furthermore, when rotating an image in Cartesian coordinates, parts of the image may be cropped to keep the same size. This was compensated for in our experiments. The performance graphs shown in Figures 2, 3, and 4 show the performance of the system under three different noisy settings. We can readily see that in all cases, the proposed system is quite comparable to the ideal exhaustive search system. In fact, under high noise conditions, the proposed technique outperforms the exhaustive method.

5. CONCLUDING REMARKS

We have presented a method for describing the rotation-invariant components of images. We showed that identifying images that only differ by a rotation can be effectively and quickly done using this technique, which has applications in Content-Based Image Retrieval and Image Registration. It is shown that the performance of the proposed method is comparable to the exhaustive search for the rotation of the two images. However, there is still room for improvements. A random matrix is used as the entanglement weight matrix \mathbf{W} , which may not contain the optimal set of weights. The amount of

discretization of the angle as a function of the size of the image must also be investigated. As the size of the image becomes larger, there needs to be more samples allocated to convert the image at larger radii into its polar representation, but exactly what the relationship is must be determined. Finally, the frequency scaling that is used to diminish the bias of the DC components of the image is empirically determined and a logarithmic scale was used; a more optimal scaling may exist, and is worthy of investigation.

6. REFERENCES

- [1] S. Zokai and G. Wolberg, "Image registration using log-polar mappings for recovery of large-scale similarity and projective transformations," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1422–1434, 2005.
- [2] F. Chaker, M.T. Bannour, and F. Ghorbel, "A complete and stable set of affine-invariant fourier descriptors," in *Image Analysis and Processing, 2003. Proceedings. 12th International Conference on*, 2003, pp. 578–581.
- [3] E. Rahtu, M. Salo, and J. Heikkila, "Affine invariant pattern recognition using multiscale autoconvolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 908–918, 2005.
- [4] S. Lazebnik, C. Schmid, and J. Ponce, "Affine-invariant local descriptors and neighborhood statistics for texture recognition," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, 2003, pp. 649–655.
- [5] Yan Ke and R. Sukthankar, "Pca-sift: a more distinctive representation for local image descriptors," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, vol. 2.
- [6] G. Carneiro and A.D. Jepson, "Multi-scale phase-based local features," in *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, 2003, vol. 1.
- [7] 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*, vol. 65, no. 1/2, pp. 43–72, 2005.
- [8] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, "Learning local affine-invariant part models for object class recognition," in *Workshop on Learning, Snowbird, Utah*, 2004.
- [9] David G. Lowe, "Distinctive image features from scale-invariant keypoints," in *International Journal of Computer Vision*, 2004.
- [10] C. Braccini and G. Gambardella, "Form-invariant linear filtering: Theory and applications," vol. 34, no. 6, pp. 1612–1628, 1986.
- [11] N. Gotze, S. Drue, and G. Hartmann, "Invariant object recognition with discriminant features based on local fast-fourier mellin transform," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, Barcelona, 2000, vol. 1, pp. 948–951.