ROTATION-INVARIANT OBJECT DETECTION USING COMPLEX MATCHED FILTERS AND SECOND ORDER VECTOR FIELDS

Mihails Pudzs, Modris Greitans

Institute of Electronics and Computer Science, Riga, Latvia

ABSTRACT

In this paper we introduce two concepts: second order vector fields that describe line-like objects in images and rotation-invariant Complex Matched Filter kernels that can be used to detect object with almost any complexity. We present the theoretical grounds for kernel derivation, object matching using sets of subresponses, object's rotation angle and active area determination. The work of the proposed algorithms is demonstrated on images of an occluded and rotated object.

Index Terms— image processing, matched filters, vector fields, angular invariance, object detection

1. INTRODUCTION

Reliable detection and recognition of objects in images are challenging tasks for a computer, since in real applications it is impossible to find two identical images of the same object, but they have to be identified as being same. These image variations include rotation, scaling [1], change in lighting conditions, or more complex ones, like pose/silhouette change. However, many applications involve images of objects of the same size, rotated at arbitrary angles, for example: automated microorganism counting/classification tasks in bio-medical imaging [2]; complex objects might also be detected by their parts, for which it might be assumed that all deformations, other than planar rotation, are negligible.

1.1. Related Work

Most of the known rotation-invariant object detection algorithms can be divided into two categories [1]: a) straightforward − working with image pixels in spatial domain, sometimes compensating rotation before detection [3], [4], b) using transformations like Radon [1], Radon-Fourier [5], Wavelet [6], HOG [7], Circular Fourier-HOG [2], with rotation-invariant properties. Latter (HOG and Circular Fourier-HOG), as well as "point-features", like SIFT [8], Edge Profile Clusters [9], etc. are also used as local image transformations − they exploit rotation invariance locally, while continue working with image in the spatial domain.

The equivariant filters, used in [10], have similar filter kernels to Generalized Complex Matched Filters (GCMF) [11] and the kernel equivariance is similar to idea of complex harmonics with opposite-phases to be built into filter kernel (subsection 2.3) to compensate for angular rotations; however, the image processing and feature comparison is different.

1.2. Our Research and its Goals

Our previous work was concerned with palm feature extraction from infrared images using Complex Matched Filters like GCMF [11], Non-Halo CMF (NH-CMF) [12] or Fast-NH-CMF [13]. All the filters are rotation covariant, but require extracted objects to be members of *line-like object* (LLO) class, which is true for palm vessels and ridges. However, further investigation shows that application of the CMF procedure twice allows to extract *almost any*[1] arbitrary object preserving rotation covariance. Our research is divided into subsequent studies. This work lays theoretical grounds for rotation invariant object detection that uses CMF and second order vector fields (2Φ), and in particular − shows that it is possible to acquire True Positive (TP) matches of arbitrarily rotated objects even in cases of defects, such as occlusions/missing parts. In some cases these defects may occupy upto half of the image area of detectable object. Further work will deal will all other possible cases like True Negatives, and suppression of False Positive (FP) responses and misses. Therefore, it is yet impossible to evaluate the performance of proposed algorithm in terms of ROC or TP/FP curves and we mostly focus our attention on theoretical derivations.

2. BASIC THEORY − LINEAR CASE

2.1. 2Φ fields − their acquisition and some properties

Second order vector fields 2Φ (or 2PHI) are *2D vectorial fields* that are generated by processing the original image $f(x, y)$ by *some acquisition algorithm* and describe this image, such that each element $\vec{c}(x, y)$ of 2Φ such that:

- its magnitude $|\vec{c}(x, y)|$ represent the similarity of neighborhood area of pixel $f(x, y)$ to LLO,
- its angle $Arg[\vec{c}(x, y)]$ represent the doubled angle of LLO, found in the neighborhood area of pixel $f(x, y)$.

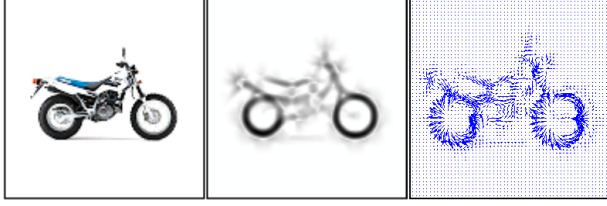[1]problematic objects will be mentioned in subsection 2.5.

Fig. 1. 2Φ illustration: original image (left), $|\underline{c}(x,y)|$ (middle), $\underline{c}(x,y)$ displayed as vectors (right); Fast-NH-CMF used

Fig. 2. An image of letter "F" in 2Φ field − unrotated (left) and rotated by an angle φ (right)

For convenience we will also refer to 2Φ fields as being *complex fields* and denote elements of 2Φ fields with an *underscore* $(\vec{c}(x,y) \equiv \underline{c}(x,y))$, because of extensively used complex algebra operations. Properties of 2Φ include: unique representation of LLO, either full linearity, or semi-linearity (with respect to LLO, being extracted by an *acquisition algorithm*). At the moment, techniques that are known for us to acquire 2Φ fields (like shown in Fig. 1) are:

- analytically − by implicit differentiation,
- second order GCMF − fully linear with Halo effect,
- NH-CMF, Fast-NH-CMF − semi-linear, no Halo effect.

When using CMF's to acquire 2Φ fields, last step of decreasing angle by half must be omitted. A *complex convolution* of two 2Φ fields $\underline{g}(x_0, y_0) = \underline{c}(x,y) \circledast \underline{M}(x,y)$ is defined by:

$$\underline{g}(x_0, y_0) = \iint_M \underline{c}(x,y) \cdot \underline{M}(x - x_0, y - y_0)dxdy, \quad (1)$$

where $\underline{c}(x,y)$ is called an "input image in 2Φ" (an example is shown in Fig. 1, right) and $\underline{M}(x,y)$ − "filter kernel in 2Φ". Since we mostly deal with 2Φ fields, we further omit specifying it every time. Complex convolution is linear operation and it can be performed using four convolution operations like so (note: (x,y) are omitted, $j = \sqrt{-1}$):

$$\underline{c} \circledast \underline{M} = (Re[\underline{c}] \otimes Re[\underline{M}] - Im[\underline{M}] \otimes Im[\underline{c}]) + \\ + j \cdot (Re[\underline{c}] \otimes Im[\underline{M}] + Im[\underline{c}] \otimes Re[\underline{M}]). \quad (2)$$

2.2. Setup for kernel derivation

Suppose we have a shape to detect in 2Φ field, Fig. 2, left. We can arbitrarily choose its rotation center; here it's shown by coordinate line intersection point. We then find minimum r and maximum R radius of where the shape pixels appear when rotating at all possible angles. The derived kernel pixels will localize only within the ring of radial distances of $[r; R]$.

Since *complex convolution* is linear operation, we can calculate $\underline{g}(0,0)$ using (1) by parts: *by subdividing whole kernel's ring of radial distances $[r; R]$ into a sum of smaller subrings, each only one pixel thick.* For example, we further observe a subring with mean radius r_0 and thickness of 1 pixel. At initial derivation state, it contains only a portion of shape's 2Φ vectors, say, $\underline{A}$ and $\underline{B}$, at angles α and β, accordingly (Fig.
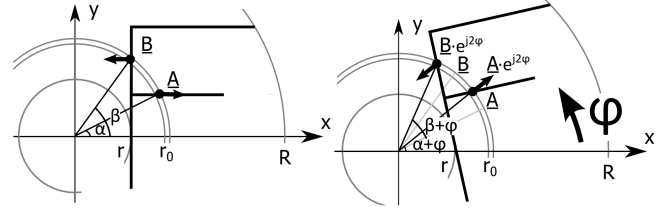
2, left). When shape is rotated around the chosen center by some angle φ (Fig. 2, right), all angles increase by the same amount, and become $\alpha + \varphi$ and $\beta + \varphi$, but since the tangents to all LLO change as well, 2Φ representation of previously known pixels become multiplied by doubled angle shift coefficient $e^{j2\varphi}$ and previously known 2Φ vectors become $\underline{A} \cdot e^{j2\varphi}$ and $\underline{B} \cdot e^{j2\varphi}$.

2.3. Derivation of M_2 kernel for object detection

Object detection M_2 kernel is defined for 2Φ field in polar coordinates by the following identity:

$$\underline{M_2}(\rho, \theta) = \underline{D}(\rho) \cdot e^{-j2\theta}, \quad (3)$$

where $\underline{D}(\rho)$ is yet unknown complex function. We derive the conclusions only for a subring of radius r_0, and using linearity equivalent conclusions might be drawn for any other subring of any other radius within the kernel area. When complex-convolving unrotated image $\underline{c}(\rho, \theta)$ (Fig. 2, left) and kernel $\underline{M_2}(\rho, \theta)$ with yet unknown $\underline{D}(r_0)$, the result at the center of the object is:

$$\underline{g}(0,0)_{r_0} = \underline{D}(r_0)e^{-j2\alpha}\underline{A} \cdot r_0 + \underline{D}(r_0)e^{-j2\beta}\underline{B} \cdot r_0. \quad (4)$$

Since all the values, except $\underline{D}(r_0)$ and $\underline{g}(0,0)_{r_0}$ are known, the equation (4) can be solved for $\underline{D}(r_0)$ as:

$$\underline{D}(r_0) = \frac{\underline{g}(0,0)_{r_0}}{r_0 \cdot (e^{-j2\alpha}\underline{A} + e^{-j2\beta}\underline{B})}, \quad (5)$$

where $\underline{g}(0,0)_{r_0}$ is the filter subring response *we demand* at this location. When the object is rotated (Fig. 2, right), complex convolution at the center yields same result:

$$\underline{g}(0,0)_{r_0}^{+\varphi} = \\ = r_0 \cdot \left(\underline{D}(r_0)e^{-j2(\alpha+\varphi)}\underline{A}e^{j2\varphi} + \underline{D}(r_0)e^{-j2(\beta+\varphi)}\underline{B}e^{j2\varphi} \right) \quad (6) \\ = r_0 \cdot \underline{D}(r_0) \cdot \left(e^{-j2\alpha}\underline{A} + e^{-j2\beta}\underline{B} \right) = \underline{g}(0,0)_{r_0}.$$

Notice how filter subring response $\underline{g}(0,0)_{r_0}$ is indifferent to rotation angle φ (*rotation invariance*). In general, for any ρ-th subring with radius ρ and $N(\rho)$ pixels in it, $\underline{D}(\rho)$ can be calculated as follows:

$$\underline{D}(\rho) = \frac{\underline{g}(0,0)_\rho}{\rho \cdot \sum_{i=1}^{N(\rho)} \underline{P}_i(\rho) \cdot e^{-j2 \cdot \theta_i(\rho)}}, \quad (7)$$

where $\underline{P}_i(\rho)$ and $\theta_i(\rho)$ are i-th pixel's value in 2Φ and angle in polar coordinates. If all the $\underline{g}(0,0)_\rho$ are chosen as positive real numbers and:

$$\int_r^R g(0,0)_\rho \, d\rho = 1, \tag{8}$$

then because of the superposition M_2 filter must always react to the same object as "1", despite object rotation. By specifying different values of $g(0,0)_\rho$ for different ρ you can specify which radius' details are more important than others. Sometimes the processed images may have different contrast and $\underline{g}(x,y) = \underline{c}(x,y) \underline{\otimes} M_2(x,y)$ may deviate from chosen "1", but *in theory it shall remain a positive real number*, which is convenient to test, for example, using this function (with imaginary part penalty coefficient α_g chosen as high as you like):

$$g(x,y) = \frac{0.5 \cdot \left(Re\left[\underline{g}(x,y)\right] + \left| Re\left[\underline{g}(x,y)\right] \right| \right)}{1 + \alpha_g \cdot \left| Im\left[\underline{g}(x,y)\right] \right|}. \tag{9}$$

2.4. Derivation of M_1 kernel for object angle estimation

Object angle estimation kernel M_1 is defined for 2Φ field in polar coordinates as follows:

$$\underline{M_1}(\rho,\theta) = \underline{E}(\rho) \cdot e^{-j1\theta}, \tag{10}$$

where $\underline{E}(\rho)$ is yet unknown complex function. Following exactly the same steps as in subsection 2.3, one can proof, that for rotated shape M_1 kernel responds as:

$$\underline{g}(0,0)_{r_0}^{+\varphi} = \underline{g}(0,0)_{r_0} \cdot e^{j\varphi}, \tag{11}$$

and that in general, for any ρ-th subring with radius ρ and $N(\rho)$ pixels in it $\underline{E}(\rho)$ can be calculated as follows:

$$\underline{E}(\rho) = \frac{\underline{g}(0,0)_\rho}{\rho \cdot \sum_{i=1}^{N(\rho)} \underline{P}_i(\rho) \cdot e^{-j \cdot \theta_i(\rho)}}. \tag{12}$$

If all the $\underline{g}(0,0)_\rho$ are chosen as positive real numbers and condition (8) is met, then because of superposition M_1 filter must always react to the same object as $e^{j\varphi}$, so at the center of an object $Arg[\underline{c}(x,y) \underline{\otimes} \underline{M_1}(x,y)] = \varphi$ (*rotation covariance*).

2.5. Known Problems for Linear Case

Both $- M_2$ and M_1 kernels exist only when the respective denominators in (3) and (10) are non-zero; and the closer the value of the denominator is to zero, the more *unstable* is the filter behavior (e.g. for noise and distortions). When constructing a kernel, subrings that fail to satisfy the non-zero denominator condition, as well as unstable subrings, might be ignored to improve overall filter stability. Known objects that fail to satisfy (10) are symmetric: circles, crosses, etc. $-$ for them it is impossible to determine their rotation angle.
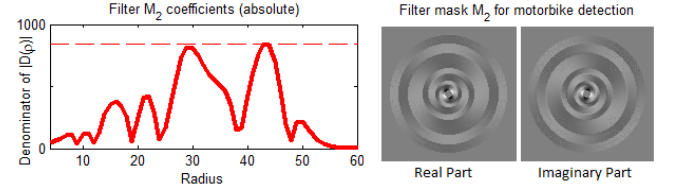

Fig. 3. M_2 kernel denominator coefficients (absolute values) (left) and M_2 kernel used for detection of bike (from Fig. 1) (right)
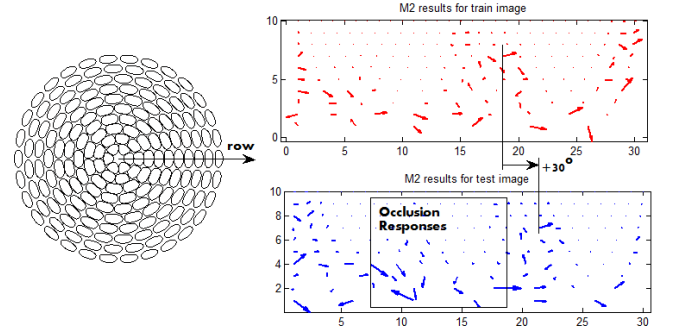

Fig. 4. Kernel segmentation mesh (left) and subresponses it produces for original image from Fig. 1 and the same occluded image, rotated by 30° counter clockwise (right)

With an object center placed in the middle of image in Fig. 1 (left) we are now able to generate kernel denominator coefficients for (7) (shown in Fig. 3, left)) and (12), as well as to generate kernel itself. M_2 kernel for angle invariant bike (from Fig. 1, left)) detection is shown in Fig. 3 (right). The main issue of linear approach is that *the kernel occupies more area than object itself* and *is prone to neighbor objects, occlusions and missing parts*, thus producing false positives and misses.

3. NON-LINEAR ENHANCEMENTS

To deal with mentioned issue we break kernel into segments (a single filter response turns into a set of subresponses), and we compare different segment subresponses as sets (to ignore occluded regions).

3.1. Kernel Segmentation

Kernel is segmented into N_ρ radial parts and for each radial parts into $N_\theta(\rho)$ angular parts. Full kernel mesh used in our experiments can be seen in Fig. 4 (left). In Fig. 4 (right) each subset of subresponses, belonging to the same radius is called "row", and we further compare subresponses row-to-row. As you can see, most of the subresponses from the training image are still identifiable (only shifted) even when image is rotated because of the rotation invariant property of

the M_2 filter. Occluded or missing parts produce the different subresponses that must be rejected/ignored. Training image also provides an *object active area mask* $m(\rho)$ where training image subresponses are above a certain threshold $-$ it is the only area where the comparison of subresponses between the training and test images take place.

3.2. Vector Row Comparison

Given a ρ-th row of subresponses (vectors) $a(\rho)$ we are going to define and use some special operations on them. A *circular shift* of row by n cells to the right will be written as $a(\rho)_n$. We will also use *element-wise operations*, like subtraction, multiplication "$*$" and absolute value calculation $abs[]$. The last operation to mention is a summation of row elements $sum[]$.

3.2.1. Discrete Comparison

For each row $\rho = 1...N_\rho$, we observe every possible shift of the training subresponse row $Tr(\rho)_\Theta$, $\Theta = 0, ...N_\theta(\rho) - 1$, and compare it with non-shifted test image subresponse row $Test(\rho)$ within the allowed area $m(\rho)_\Theta$ by calculating error:

$$\Delta(\rho, \Theta) = sum\left[abs\left[Test(\rho) - Tr(\rho)_\Theta\right] * m(\rho)_\Theta\right]. \quad (13)$$

3.2.2. Comparison Using Linear Interpolation

The comparison method stays the same, it is only supposed that object can be rotated by a fraction of cells that kernels was originally divided into (Fig. 4, left). Simplest approach to overcome this difficulty is to use linear interpolation and instead of comparing $Tr(\rho)_\Theta$ with the test image, use interpolated trained row:

$$Tr(\rho)_{[\Theta+k]} = (1-k) \cdot Tr(\rho)_\Theta + k \cdot Tr(\rho)_{\Theta+1}, \quad (14)$$

where optimal value for $k \in [0, 1]$ can be found using *Least Square Error* $\Delta(\rho, [\Theta + k])$ *minimization*. In both cases, the position either $\Theta(\rho)$, or $[\Theta + k](\rho)$ with the smallest Δ, as well as the difference in rows is saved, because it provides the most similar subresponse row alignment. $[\Theta + k](\rho)$ is a more precise circular shift that can be determined in the row comparison stage, and it is not so dependent of kernel segment sizes. Furthermore, knowing the angular length of each segment in each row $\delta_\theta(\rho)$, the angular shift of the row can be then calculated as:

$$\varphi(\rho) = [\Theta + k](\rho) \cdot \delta_\theta(\rho). \quad (15)$$

3.3. Accurate object angle estimation

Statistically, object rotation angle might be estimated using histogram of determined $\varphi(\rho)$ values over all analyzed radiuses ρ. However, traditional approach to drawing histograms with dividing angular interval $[0, 360)$ into N_φ bins is not precise enough, because it produces round-offs
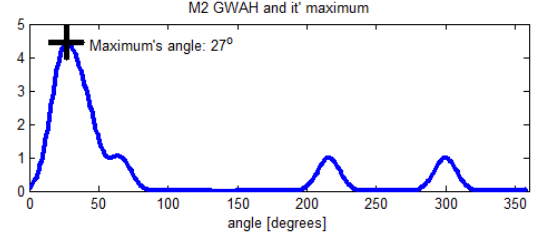

Fig. 5. GWAH for heavily (about 50%) occluded and rotated by $+30^o$ object from Fig. 1

when assigning values into a specific bin. Therefore, we use Gaussian-like function Weighted Angular Histogram (GWAH) for object angle determination, where each peak with angle $\varphi(\rho)$ adds a Gaussian-like function, defined as:

$$G(\varphi - \varphi(\rho)) = e^{-\frac{(\varphi - \varphi(\rho))^2}{2\sigma^2}}, \quad (16)$$

to a continuous histogram. Constant σ is chosen to be $^1/_3$-rd of a cell's angular width, so the function overlaps the region of 2 adjacent cells with its 99% of energy. An example of GWAH is shown in Fig. 5: despite more than 50% missed guesses of the comparison algorithm, the determined angle of an object is still very accurate: 27°.

3.4. Final Comparison

Once the object's angle is determined as Mode of GWAH: $\varphi_{obj} = Mode(GWAH(\varphi(\rho)))$, the process described by (15) can be reversed to acquire correct $[\Theta + k]_{\varphi_{obj}}(\rho)$ for every ρ. Then, the comparison is repeated by calculating $\Delta(\rho)_{[\Theta+k]_{\varphi_{obj}}(\rho)}$ and thresholding it. Subresponses that didn't pass the thresholding stage are probably distorted: represent missing or occluded parts. We can also calculate M_2 and M_1 responses $-$ total, and for each subring separately.

3.5. Summary $-$ Process Flow

To sum it up, when processing an image (that is, for every coordinate), to acquire a True Positive:
1. image region is complex-multiplied with M_2 and, if needed, M_1 kernel segments,
2. discrete and interpolated comparison of subresponses is performed,
3. iff different rows disagree on object rotation angle, GWAH is analyzed,
4. subresponse rows are compared, object's active area and filter responses are calculated,
5. finally, (9) is used to test if the object is found.

4. EXPERIMENTAL RESULTS

Figure 6 demonstrates the performance of described algorithm on a distorted rotated image from Fig. 1. A bright area
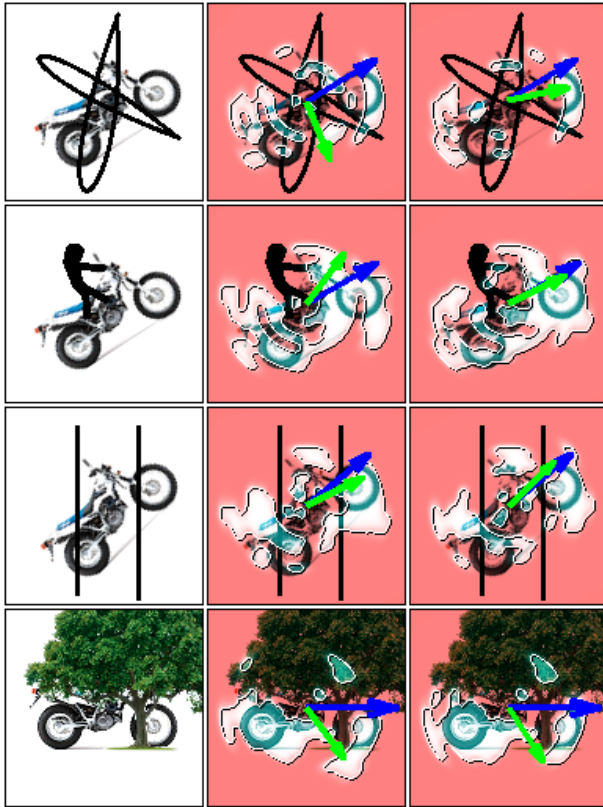
Fig. 6. Test images analyzed by described method, from left to right: original image, result before GWAH analysis, result after GWAH analysis

denotes matched subresponse regions; longer (blue) arrow represents the angle of the object, determined by GWAH, shorter (green) arrow − angle, determined by M_1 filter. Since most of image area was ignored (and not shown as bright), the results may disagree. The repeated set comparison (after φ_{obj} estimation) improves the M_1 angle estimate. To check whether an object is found (9) can be used, but it must be noted that *filter response is proportional to the matched area*, which is a known value. Computational cost of image convolution with segmented kernel is not greater than an image convolution with a kernel itself. Step 3.2.1 may be omitted at all, because acquiring either $k < 0$, or $k > 1$ means a discrete-shifted row match. A choice of threshold for row difference is still a subject of study, as well as futher filter developments.

REFERENCES

[1] Tomasz Arodz, "Invariant object recognition using radon-based transform.," *Computers and Artificial Intelligence*, vol. 24, no. 2, pp. 183–199, 2005.

[2] H. Skibbe and M. Reisert, "Circular fourier-hog features for rotation invariant object detection in biomed-ical images," in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, May 2012, pp. 450–453.

[3] H.A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*, Jun 1998, pp. 38–44.

[4] M. Villamizar, F. Moreno-Noguer, J. Andrade-Cetto, and Alberto Sanfeliu, "Efficient rotation invariant object detection using boosted random ferns," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, June 2010, pp. 1038–1045.

[5] Wenxing Ye, Christopher Paulson, Dapeng Oliver Wu, and Jian Li, "A rotation-invariant transform for target detection in sar images," 2008, vol. 6970, pp. 69700W–69700W–11.

[6] Du-Ming Tsai and Cheng-Huei Chiang, "Rotation-invariant pattern matching using wavelet decomposition," *Pattern Recogn. Lett.*, vol. 23, no. 1-3, pp. 191–201, Jan. 2002.

[7] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, June 2005, vol. 1, pp. 886–893 vol. 1.

[8] David G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision - Volume 2*, Washington, DC, USA, 1999, ICCV '99, pp. 1150–, IEEE Computer Society.

[9] N. G. Kingsbury R. Anderson and J. Fauqueur, "Rotation-invariant object recognition using edge-profile clusters," in *Proceedings of European Conference on Signal Processing*, September 2006.

[10] Qing Wang O. Ronneberger, W. Driever and Kun Liu, "2d/3d rotation-invariant detection using equivariant filters and kernel weighted mapping," *2013 IEEE Conference on Computer Vision and Pattern Recognition*, vol. 0, pp. 917–924, 2012.

[11] M. Pudzs, M. Greitans, and R. Fuksis, "Generalized complex 2d matched filtering for local regular line-like feature detection," in *European Signal Processing Conference*, 2011, pp. 41–45.

[12] M. Pudzs, M. Greitans, and R. Fuksis, "Complex 2d matched filtering without halo artifacts," in *Systems, Signals and Image Processing (IWSSIP), 2011 18th International Conference on*, June 2011, pp. 1–4.

[13] M. Pudzs, R. Fuksis, R. Ruskuls, and M. Greitans, "Biometric authentication apparatus and biometric authentication method," *WO 2013/147574 A1*, Oct 2013.