QUESTIONING HU'S INVARIANTS
Bad or Good Enough?

Diego Martinoia

Department of Electronics and Information, Politecnico di Milano, P.za Leonardo da Vinci 32, 20133, Milano (MI), Italy

Keywords: Gesture Recognition, Invariants, Moments, Shape.

Abstract: Despite Hu's invariants were proven not to be independent nor complete long time ago, their use in computer vision applications is still broad, mainly because of their diffusion among common CV libraries and ease of use by inexperienced users. In this paper I want to investigate whether, given their mathematical flaws, they are nevertheless good enough to justify such a wide diffusion, also considering that more sophisticated tools have been developed over the years.
In order to do this, I am going to test the robustness of Hu's invariants in a comparative way against the more modern wavelet invariants, in a hand gesture recognition application. Finally, I am going to discuss, basing my considerations on the experimental data, whether Hu's invariants are still a viable option for small scale, amateurish applications, or if the time has come to abandon them for more effective solutions.

1 INTRODUCTION

Image invariants can be defined as characteristic values of an image, chosen to be invariant to some kind of distortions, usually a combination of translation, rotation and scaling. Albeit invariants suffer of various weaknesses, such as the inability to be used to match occluded objects, they are still considered one of the main benchmark references for all other shape descriptors. Over the course of the years, moment invariants have become one of the most famous and common tools in shape recognition. Moment invariants find their roots in the theory of algebraic invariants, and they have been used for the first time in pattern recognition from Hu (Hu, 1962), who developed the very same set of invariants whose robustness I am going to examine in this paper. After that, many authors have either improved and generalized Hu's work, or applied his results to many different application fields, from satellite images (Wong and Hall, 1978) to hand writing recognition (Flusser and Suk, 1994). Flusser (Flusser, 2000), probably inspired by the previous work of Reiss (Reiss, 1991), developed a method to derive a complete set of independent invariants for every image, and proved mathematically that Hu's invariants are not independent, nor complete. Nevertheless, their use in the field of computer vision is still broad, mainly because of their ease of use, being implemented in many of the available libraries, such as the widely used OpenCV library (WillowGarage, 2011). On the other hand, we see very

few implementations, in those very same libraries, of more recent and effective sets of invariants, such as Hermit-Gaussian's, Li's, Zernike's or wavelet invariants. Therefore, a non-expert user of those libraries, who might be willing to implement his own ideas, will probably be using, consciously or not, Hu's set rather than other ones. My aim is to test whether this is a viable choice for small scale, amateurish applications, given the low complexity and the ease of use of Hu's set, or if the tradeoff is too extreme, and it is time, after 50 years since their invention, to start making pressure on the CV libraries maintainers to have them deprecate Hu's set, placing it side by side with the more recent tools. To address this question, I am going to compare the performance of Hu's set with the one of a subset of the wavelet invariants, in a configuration where the performance of the acquisition device is comparable to those of amateurish devices, i.e. low resolution and low frame rate, and the application is hand gesture recognition.

2 THEORETICAL DESCRIPTION

In this section I am going to give to the reader some theoretical background about the general procedure in order to obtain a set of invariant features, immune to Translation, Rotation and Scaling (TRS) modifications, as well as describe Hu's and wavelet sets more specifically.

Let f(x,y) by a 2-D binary image object in the (x,y)

Cartesian coordinate space, while its corresponding polar representation is $f(\rho, \theta)$. We can achieve TS invariance by using two new coordinates, that are defined as:

$$x' = \frac{x - X_0}{\alpha}, y' = \frac{y - Y_0}{\alpha}$$

where (X_0, Y_0) are the coordinates of the center of mass of the object of interest, and α is the square root of the ratio between the object's size and its expected size. From now on, I am going to refer to $f(x, y)$ and $f(\rho, \theta)$ as the TS normalized version of the original image, respectively in Cartesian and polar coordinates.

R normalization is achieved as follows: Let

$$F_{pq} = \int \int f(\rho, \theta) g_p(\rho) e^{jq\theta} \rho \, d\rho \, d\theta \qquad (1)$$

being $g_p(\rho)$ a function of the radial variable ρ, and p and q are integer parameters. The demonstration of the rotation invariance of $\left\| F_{pq} \right\|$ is achieved as follows: if the image $f(\rho, \theta)$ is rotated of an angle β, its corresponding moment becomes $F_{pq}^{\beta Rot.} = F_{pq} e^{jq\beta}$, that as we know doesn't affect the value of the norm of the vector. Thus, $\left\| F_{pq} \right\|$ is a rotation invariant.

As shown by Shen (Shen, 1999), we can derive that Hu's moments, Li's moments and Zernike moments are all special cases of (1) and that the extracted features are global features.

2.1 Hu's Invariants

For the complete theoretical explanation of Hu's work, I address the reader to the original paper by Hu (Hu, 1962), and the following work by Reiss (Reiss, 1991) and Flusser (Flusser, 2000). The set of the 7 original Hu's moment invariants can be directly derived from the geometrical center moments:

$$\phi_1 = \eta_{20} + \eta_{02}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)$$

$$\phi_6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2) + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$$

$$\phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 -$$

$$3(\eta_{21} + \eta_{03})^2) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2)$$

where

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{(1 + \frac{p+q}{2})}}$$

and

$$\mu_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - X_0)^p (y - Y_0)^q f(x, y) \, dx \, dy$$

is the central moment of the function $f(x, y)$. One of the main weaknesses of the set is its explosive complexity, and thus its usage is almost always limited to these 7 invariants, albeit higher order moments have been constructed, for example by Li (Li, 1992) and Wong et Al (Wong et al., 1995), respectively up to order nine and five. Flusser (Flusser, 2000) proved that the above mentioned set is not independent: let c_{pq} be the pq-order complex moment in polar coordinates denoted by:

$$c_{pq} = \int_0^{+\infty} \int_0^{2\pi} \rho^{p+q+1} e^{i(p-q)\theta} f(\rho, \theta) \, d\rho \, d\theta$$

Under this representation, Hu's invariants can easily be rewritten as:

$$\phi_1 = c_{11},$$
$$\phi_2 = c_{20}c_{02},$$
$$\phi_3 = c_{30}c_{03},$$
$$\phi_4 = c_{21}c_{12},$$
$$\phi_5 = \Re(c_{30}c_{12}^3),$$
$$\phi_6 = \Re(c_{20}c_{12}^2),$$
$$\phi_7 = \Im(c_{30}c_{12}^3)$$

From this representation we can see that:

$$\phi_3 = c_{30}c_{03} = \frac{c_{03}c_{21}^3 c_{30}c_{12}^3}{(c_{21}c_{12})^3} = \frac{\phi_5^2 + \phi_7^2}{\phi_4^3}$$

and therefore the invariants are not independent.

2.2 Wavelet Invariants

A detailed explanation of the wavelet invariants can be found in Shen (Shen, 1999), while some of their applications for image matching can be found in Zhang et al (Zhang et al., 2009).

Wavelet Transform is a method for obtaining localized analysis, but unlike the traditional short-time Fourier Transform, Wavelet Transform can provide both time and frequency localization. While dealing with the wavelet based approach, we are going to treat $g_p(\rho)$ from (1) as the wavelet mother function, and consider the wavelet family:

$$\psi^{a,b}(\rho) = \frac{1}{\sqrt{a}} \psi\left(\frac{\rho - b}{a}\right) \qquad (2)$$

where $a, b \in R_+$ are, respectively, the dilatation parameter and the shifting parameter. In the experiment I am going to use the cubic B-spline in Gaussian approximation form (Unser et al., 1992), given by:

$$\psi(\rho) = \frac{4a^{n+1}}{\sqrt{2\pi(n+1)}} \sigma_w \cos(2\pi f_0(2\rho - 1)) e^{-\frac{(2r-1)^2}{2\sigma_w^2(n+1)}}$$

where $n = 3, a = 0.697066, f_0 = 0.409177, \sigma_w^2 = 0.561145$. This choice was based on the properties illustrated in (Ahuja et al., 2005).

The values of the a and b parameters in (2) are usually discrete, and discretization is achieved by choosing $a = a_0^m$ with m integer and $a_0 \neq 1$ while b is discretized by taking the positive and negative multiples of $b_0 a_0 m$, with b_0 chosen so to cover the whole domain of $\psi((\rho - b)/a)$ for different values of m or, in other words:

$$a = a_0^m, b = n b_0 a_0 m$$

with n,m integer. But since the image size ha been normalized in the domain $\rho \leq 1$, one can set a_0 and b_0 both to 0.5, and restrict the domain for m and n as:

$$m = 0, 1, 2, 3; n = 0, 1, ..., 2^{m+1};$$

in this way, we obtain a simplified form of the wavelet, defined along a radial axis in any orientation:

$$\psi_{m,n} = 2^{\frac{m}{2}} \psi(2^m \rho - 0.5n)$$

Now, if we let this function sweep through all the angular rotation in the moment computation, it will extract global or local features according to the values of m and n.

The general form of a wavelet invariant is given by:

$$\left\| F_{m,n,q}^{\text{wavelet}} \right\| = \left\| \int \int f(\rho, \theta) \psi_{m,n}(\rho) e^{jq\theta} \rho d\rho d\theta \right\|$$

where $\psi_{m,n}(\rho)$ replaces $g_p(\rho)$ in (1), $m = 0, 1, 2, 3; n = 0, 1, ..., 2^{m+1}; q = 0, 1, 2, 3$. Therefore in our case we have 136 possible wavelet invariants, while Hu's are only 7. In order to make the comparison fairer, one has to choose a reduced set of the wavelet invariants.

2.3 Choosing the Invariants

All of Hu's invariants were kept, due to their limited number. Besides, the purpose of this test is to verify the set's robustness as it is, including the dependent values. For the wavelet invariants, the first reduction I have chosen is the restriction of the domain of the q parameter to the single 0 value. Albeit this might seem like a drastic choice, we must remember that we want to go from 136 dimensions to 7, and no matter

what reduction we choose, we are going to lose a lot of information in the process. At least, this choice allows to avoid working with complex numbers, reducing the computational load on that part, and allowing to increase the resolution of the dρ and dθ values of the discretization of the polar integral, thus reducing numerical instabilities.

After some exploration of the parameters space, I found out that the invariants with lower variance over the training set in the 34 remaining options were (for each of them, $q=0$ and therefore I report only the values of m and n): $\left\| F_{0,0}^{\text{wavelet}} \right\|$, $\left\| F_{1,3}^{\text{wavelet}} \right\|$, $\left\| F_{2,0}^{\text{wavelet}} \right\|$, $\left\| F_{2,1}^{\text{wavelet}} \right\|$, $\left\| F_{2,2}^{\text{wavelet}} \right\|$, $\left\| F_{3,0}^{\text{wavelet}} \right\|$, $\left\| F_{3,1}^{\text{wavelet}} \right\|$, which are distributed over all the values of m.

3 EXPERIMENTAL SETUP

The acquisition device I am going to use in this experiment is the Mesa Imaging SwissRanger SR4000 Time of Flight camera (Imaging, 2011). It acquires a black and white, 176x144 resolution, 16bit depth image, providing reliable information about the distance of each pixel from the camera when the pixel is located in the $30 - 300$ cm distance range. The camera frame-rate used is 20 FPS.

Since the design of an effective hand-detection algorithm is beyond the purpose of this paper, the hand segmentation is performed using distance information, in order to create a scanning layer and obtain a binary image. Eventual remaining holes and noise are then subjected to morphological transformations for filling and filtering. A similar approach for segmentation has already been used for touch-less interfaces, with excellent results (Soutschek et al., 2008).

At the start of the experiment, I generate a number N of agents, each one storing the $K = 7$ invariant values of a specific gesture out of the M gestures in the database, independently perturbed by a uniformly distributed mutation, which magnitude is in the $\pm V\%$ of the reference value. Each agent stores the signature of only one gesture, therefore there are at least $\lfloor N/M \rfloor$ agents evaluating each gesture. I am going to refer to a set of agents evaluating the same gesture as "committee".

The evaluation is performed using the following procedure: for each frame, for each agent, I am going to calculate the average percentage variation of the logarithms of the invariants of the image, compared to those of the reference, which is the "distance" between the stored reference and the current image. If the result is bigger than a given tolerance T, then the shape is considered as not matching the shown ges-

ture, yes otherwise. This very same decision policy is used even in the OpenCV libraries implementation of Hu's invariants (WillowGarage, 2011). At the end of this preliminary evaluation, I am going to calculate the number of agents that recognized their gesture in the current frame, but only when a minimum percentage of agreement was reached in their committee, assign at each gesture a probability accordingly and randomly extract the output. If no committee reaches *P%* recognition, then no gesture is recognized in that frame.

4 EXPERIMENTAL RESULTS

User dependency is a very big issue in gesture recognition, as proven by the success in identification applications (Schmidt et al., 2010), and the evaluation of its effects is clearly outside the scope of this paper. Therefore, I am going to use the same users for both training and testing. Albeit the gesture database is very limited, consisting only of 5 gesture performed with both hands, the experiment should be able to give us some useful material for discussion. I performed 3 different experiments, testing the performances of the sets over still images, moving images, and still images with added noise respectively.

Due to the users' inability to stay perfectly still in posture, slight changes in the perceived image occurred. In order to avoid setting the value of *T* too high to compensate, testing was performed also over a "Noise" gesture, i.e. a gesture not matching any of the database ones, but which looked similar to many of them under different perspective angles, shown rotating on all possible axis.

Rock   *Paper*   *OK*   *Stop*   *Pointing*   *Noise*
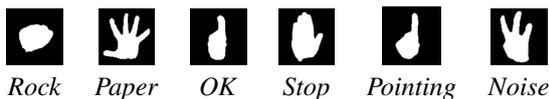
I report only the best results, over all the parameters combinations tested. In all cases, training was performed using 15 seconds long recordings of the gesture, recorded in a static pose, while testing was done using 10 seconds long recordings (which are long times, considering the application). Training consisted in calculating the invariants of each frame of the training sample, and then averaging the output only when the value was not further than $\pm50\%$ of the median value, in order to filter out training noise. Due to the comparative nature of this study, the very same data-set for training and experiment were used for both Hu's and wavelet approaches.

Table 1: Static gestures; Hu's invariants; N=300; M=5; V=5%; T=7.5%; P=75%.

Gesture	Correct	Opposite	Total	Wrong
R. Rock	96.9	0.0	96.9	0.0
R. Paper	89.3	0.0	89.3	0.0
R. OK	97.2	2.1	99.3	0.0
R. Stop	93.7	0.0	93.7	0.0
R. Pointing	95.7	0.0	95.7	0.0
L. Rock	92.2	0.0	92.2	0.0
L. Paper	97.3	0.0	97.3	0.0
L. OK	99.2	0.0	99.2	0.0
L. Stop	73.7	25.2	99.3	0.0
L. Pointing	97.1	0.0	97.1	0.0
Noise	–	–	–	7.5
Average	93.2	2.8	96.0	0.8

Table 2: Static gestures; Wavelet invariants; N=300; M=5; V=10%; T=15%; P=75%.

Gesture	Correct	Opposite	Total	Wrong
R. Rock	50.4	46.5	96.9	0.8
R. Paper	83.9	0.0	83.9	5.4
R. OK	7.1	0.7	7.8	0.0
R. Stop	44.4	23.8	68.3	6.3
R. Pointing	64.0	3.6	67.6	1.4
L. Rock	53.5	42.9	96.3	1.2
L. Paper	84.6	0.0	84.6	2.0
L. OK	84.1	3.8	87.9	0.0
L. Stop	46.1	52.0	98.0	0.0
L. Pointing	20.1	0.0	20.1	2.2
Noise	–	–	–	21.5
Average	53.8	17.3	71.1	3.7

4.1 Static Experiment

All gestures were presented frontally, with the palm of the hand towards the camera. Each table represent the percentage of the ratio between the number of frames where a gesture was output, and the total number of frames. Rows represent which gesture sample was shown to the camera. "Correct" means a correct hand and gesture recognition, "Opposite" means correct gesture but wrong hand. These two columns were summed in the "Total" column, for ease of the reader. "Wrong" means a completely wrong gesture recognized, while no-output frames are not reported in this tables. The results are shown in Table 1: Hu's invariants matching ratio for static gestures is very high, with complete lack of error when a real gesture is shown to the system. In the second phase of the first experiment the test was repeated using the wavelet invariants approach. The results are shown in Table 2: as we can see, the performance of the wavelet invariants is generally lower than Hu's, which is probably due to both the discretization process for the radial integrals in such a low-resolution environment, prone

Table 3: Moving gestures; Hu's invariants; N=300; M=5; V=5%; T=7.5%; P=75%.

Gesture	Correct	Opposite	Total	Wrong
R. Rock	37.4	1.0	38.4	1.4
R. Paper	45.3	1.1	46.4	0.0
R. OK	51.0	0.0	51.0	3.5
R. Stop	69.1	0.0	69.1	0.0
R. Pointing	83.9	0.0	83.9	0.0
L. Rock	39.4	0.0	39.4	0.0
L. Paper	43.5	13.0	56.5	0.0
L. OK	43.6	12.8	56.4	0.0
L. Stop	71.3	13.1	84.4	0.0
L. Pointing	77.9	3.2	35.2	0.0
Noise	–	–	–	2.5
Average	56.2	4.4	56.1	1.1

Table 4: Moving gestures; Wavelet invariants; N=300; M=5; V=10%; T=15%; P=75%.

Gesture	Correct	Opposite	Total	Wrong
R. Rock	44.6	42.9	87.6	2.8
R. Paper	49.2	1.1	50.3	9.0
R. OK	8.1	6.6	14.7	3.6
R. Stop	17.5	12.4	29.9	25.2
R. Pointing	23.2	6.7	29.9	3.1
L. Rock	45.7	42.8	88.5	3.8
L. Paper	23.6	1.4	25.0	4.6
L. OK	9.6	6.9	16.5	3.7
L. Stop	40.8	33.6	74.3	10.
L. Pointing	12.2	6.5	18.7	1.6
Noise	–	–	–	21.5
Average	25.9	15.7	41.6	8.1

to numerical instabilities. As a matter of fact, in order to obtain these results I had to use a very high value of the tolerance T parameter, which led to a big false positive ratio. To be noted is also the inability of the wavelet set to distinguish between mirrored images (in this case left or right hand) properly. Hu's set, instead, can do this better, being the sign of ϕ_7 sensitive to mirroring.

4.2 Moving Experiment

This time, gestures in the sample were shown translating all over the camera plane, but without any rotation. Involuntary rotation and shape change happened due to users' movement, which is the cause for worse results. Results are shown in Table 3: as we can see, they are sensibly worse. Upon close inspection of the errors for the gestures that had a poorer performance ("Pointing", "OK"), I noticed that this is due to their similar outer silhouette, which is confused by the algorithm in most of the non-static situations. The results for the wavelet invariants tested over moving gestures are shown in Table 4. Wavelet invariants' performance is again generally lower than its oppo-

Table 5: 2% Salt&Pepper noise; Hu's invariants; N=300; M=5; V=5%; T=12.5%; P=75%.

Gesture	Correct	Opposite	Total	Wrong
R. Rock	56.6	6.2	62.8	0.8
R. Paper	26.8	0.0	26.8	0.0
R. OK	61.7	9.2	70.9	0.0
R. Stop	56.4	0.8	57.1	0.0
R. Pointing	63.3	0.0	63.3	0.7
L. Rock	62.9	0.4	63.3	0.4
L. Paper	47.0	0.0	47.0	0.0
L. OK	81.1	0.0	81.1	0.0
L. Stop	47.4	20.4	67.8	0.0
L. Pointing	70.5	0.0	70.5	0.0
Noise	–	–	–	2.5
Average	57.4	3.7	61.1	0.4

nent's: wavelet invariants are too sensitive to shape changes to outperform the more "stubborn" Hu's set in this setting. Another thing that needs to be pointed out is that, just like it happened with Hu's set, some gestures (for example "Rock") are much more stable than others ("OK","Pointing"). The reason for this is probably in the fact that the user can perform those gestures in a moving way with less involuntary rotation than the others, which might be an interesting consideration for a possible human interface application.

4.3 Noise Experiment

The last part of the experiment consisted in testing over the same static set as before, but including a 2% Salt&Pepper noise, i.e. negating the binary value of 2% of the pixels of the image. It is important to state that this disturbance was introduced before the morphology operations, and therefore its effect reaches the invariants extraction phase highly amplified, as shown in this sample image:

Original   *Noisy*

Results for this test are shown in Tables 5 and 6, for Hu's and wavelet invariants respectively. We can see that the average performance is deeply affected in both cases, but Hu's invariants resist better than the wavelet ones, albeit they required an increase of the T parameter, while the wavelet's did not. The reason for this is the fact that wavelet invariants are much more sensitive to the noise-induced shape differences than Hu's invariants, and are in fact used, for example, to distinguish between very similar shapes (Shen, 1999).

Table 6: 2% Salt&Pepper noise; Wavelet invariants; N=300; M=5; V=10%; T=15%; P=75%.

Gesture	Correct	Opposite	Total	Wrong
R. Rock	27.1	35.7	62.8	22.5
R. Paper	41.1	0.0	41.1	48.2
R. OK	24.8	5.1	29.8	17.8
R. Stop	31.0	17.5	48.4	25.4
R. Pointing	35.3	5.8	41.0	31.0
L. Rock	30.6	30.2	60.8	20.0
L. Paper	43.0	2.0	45.0	28.2
L. OK	50.0	8.3	58.3	9.9
L. Stop	40.8	33.6	74.3	18.4
L. Pointing	12.2	6.5	18.7	17.3
Noise	–	–	–	22.3
Average	33.6	14.4	48.0	23.6

5 CONCLUSIONS

The goal of this paper was to investigate whether Hu's moment invariants, despite their proven mathematical imprecision, are still a viable option for non-expert users and small scale applications, thanks to their simplicity. To address this question, I compared their performance with the more recent set of wavelet invariants, in a severe low-resolution environment that emulates a possible amateurish hardware scenario.

The results, albeit few in quantity, show how Hu's invariants perform greatly over still images, definitely better than the more advanced wavelet set. I have then tried to have Hu's and wavelet sets detect translating gestures: results were, obviously, much worse, but still in the 40–60% range, with Hu's invariants performing generally better than the wavelet's also in this case. Finally the test was conducted over noisy images. In both cases there is a worsening of the results of about absolute 30%, but Hu's invariants are definitely better in this case also. The reason for this is dual: firstly, the low resolution leads to problems in the discretization of the radial integral steps for the wavelet invariants calculation, and secondly the wavelet set is much more sensitive to the shape changes, intrinsic in this application, than Hu's set.

In conclusion, it is my opinion that Hu's invariants have proven their robustness under difficult conditions, and therefore they are still a great choice, especially in order to allow amateurs of the field to try and implement their ideas. Nevertheless, this doesn't mean that one must think of them as the panacea of all shape recognition problems: in other applications other approaches have proven their superiority, and CV libraries maintainers should indeed start implementing the newer tools in their collections, leaving the responsibility of picking the right tool for each problem to the developer.

REFERENCES

Ahuja, N., Lertrattanapanich, S., and Bose, N. (2005). Properties determining choice of mother wavelet. *Vision, Image and Signal Processing, IEEE Proceedings*, 152(5):659–664.

Flusser, J. (2000). On the independence of rotation moment invariants. *Pattern Recognition*, 33(9):1405–1410.

Flusser, J. and Suk, T. (1994). Affine moment invariants: a new tool for character recognition. *Pattern Recognition Letters*, 15(4):433–464.

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. In *Information Theory, IRE Transactions on*, volume 8, pages 179–187. IEEE Computer Press.

Imaging, M. (2011). Mesa imaging product description, 2011.

Li, Y. (1992). Reforming the theory of invariant moments for pattern recognition. *Pattern Recognition*, 25:723–730.

Reiss, T. (1991). The revised fundamental theorem of moment invariants. *IEEE Trans. Pattern Anal. Mach. Intell*, 13(8):830–834.

Schmidt, D., Ki Chong, M., and Gellersen, H. (2010). Handsdown: hand–contour–based user identification for interactive surfaces. *NordiCHI 2010 Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 432–441.

Shen, D. (1999). Discriminative wavelet shape descriptors for recognition of 2-d patterns. *Pattern Recognition*, 32:151–165.

Soutschek, S., Penne, J., Hornegger, J., and Kornhuber, J. (2008). 3-d gesture-based scene navigation in medical imaging applications using time-of-flight cameras. *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–6.

Unser, M., Aldroubi, A., and Eden, M. (1992). On the asymptotic convergence of b-spline wavelets to gabor functions. *IEEE Trans. Inform. Theory*, 38(2):864–872.

WillowGarage (2011). Opencv official documentation.

Wong, R. Y. and Hall, E. L. (1978). Scene matching with invariant moments. *Computer Graphics and Image Processing*, 8(1):16–24.

Wong, W., Siu, W., and Lam, K. (1995). Generation of moment invariants and their uses for character recognition. *Pattern Recognition Letters*, 16(2):115–123.

Zhang, F., Liu, S.-q., Wang, D.-b., and Guan, W. (2009). Aircraft recognition in infrared image using wavelet moment invariants. *Image and Vision Computing*, 27(4):313–318.