Localization Estimation for Autonomous Aerial Navigation by matching Images with Different Resolutions

KAMEL BENSEBAA, MAURICIO POZZOBON MARTINS
Institute for Advanced Studies (IEAV)
Brazilian General Command for Aerospace Technology (CTA)
Rodovia dos Tamoios, km 5,5 - Putim - Cep - 12.228-001
Caixa Postal 6044 - Cep - 12.228-970
São José dos Campos (SP)
BRAZIL

http://www.ieav.cta.br

Abstract: - Reliable localization is an essential component of a successful Autonomous Aerial Navigation. In order to get a precise position before navigation uncertainty becomes incorrigible, the main objective of this work is to investigate efficient algorithms for localization estimation for Autonomous Aerial Navigation by matching Images with Different Resolutions. In this work, we present an approach to localize the UAV (Unmanned Aerial Vehicle) in flight using matching images with different resolutions: a high-resolution image and a low-resolution one. This method consists of firstly to detect the landmarks or feature points in the images using the Harris corner detector. Secondly the method consists of automatic extraction of correspondence points between the first UAV video frame and a georeferengeced image. Finally we use a geometric model estimation to map the high-resolution image onto a low-resolution image, which results in position estimation of the UAV.

Key-Words: - UAV, Feature extraction, Harris detector, Ransac, Matching, Scales, Resolutions, Localization, Estimation

1 Introduction

In recent years the UAVs (Unmanned Aerial Vehicles) have become an important and highly suitable means for tasks such as: (i) remote sensing, (ii) surveillance and (iii) monitoring the terrestrial surface.

The most important objective in UAV's autonomous navigation research is to avoid human control during all flight. Its advantage is to increase the vehicle range beyond the radio frequency control.

The most common commercial navigation system for UAV (Unmanned Aerial Vehicle) is based on signal fusion between an Inertial Navigation System (INS) with an Inertial Measurement Unit (IMU) and a Global Navigation Satellite System (GNSS) [10]. Although some papers are being done with this approach, problems can occur during GNSS signal reception or with INS precision.

The methodology suggested in this work is the use of computer vision techniques to provide a robust algorithm to localize the UAV in flight. The system compares images from an onboard camera with a georeferenced image, stored previously in the system's memory to find similar features in both images, and then estimates the UAV's position through the feature coordinates extracted from the georeference image.

The computer vision system has to capture video ground images and compare them with georeferenced images in order to get a precise position before navigation uncertainty becomes incorrigible. In this paper it's considered that the aerial vehicle has at least one inertial navigation systems (INS) onboard, sending data to a hypothetical navigation system to control the UAV. So, the computer vision system must frequently calculate the position and update the INS as often and as accurately as possible.

In this work, we describe an approach to estimate the localization of the UAV using matching images with different resolutions: a high-resolution image and a low-resolution one. The high-resolution images are video frames obtained from a camera fixed to a helicopter in a low level flight, simulating the vision system of an UAV. The low-resolution images are obtained from georeferenced images. This method consists of three main steps: The first step is to detect the landmarks or feature points in the images. This is done using the Harris corner detector [6]. The second step consists of automatic extraction of correspondence points between the first UAV video

frame and a georeferenced image. Finally we use a geometric model estimation to map the high-resolution image onto a low-resolution image, which results in position estimation of the UAV. This paper is organized as follows: Section 2 gives some information about the images used in this work. Section 3 describes the algorithm with the details of each step. Sections 4 and 5 discusses the results and presents some conclusions.

2 Data Preparation

The initial task is the selection of images, which will be used in the algorithm. The scene represents the soccer stadium of São José dos Campos in São Paulo State (Brazil).

The first image is referred to the high-resolution image acquired by a camera fixed to a helicopter and presents some degradation due to interlacing process. This degradation is due to aircraft vibrations during the acquisition process. So, a pre-processing operation based on low-pass filter allows improving the original image. Fig.1 shows the degraded high-resolution image while Fig.2 displays the pre-processed image.

The second image is referred to the low-resolution image obtained from georeferenced images and represents the same scene of the high resolution image i.e. the soccer stadium of São José dos Campos in São Paulo State (Brazil). Fig.3 shows the low-resolution image.

Fig.1: Original frame image.

Fig.2: Pre-processed frame image.

Fig.3: Original georeferenced image.

3. Algorithm

The algorithm used in this work to estimate the localization of the UAV by matching images with different resolutions proceeds as follows:

- *Feature extraction* - In each image we extract salient structures (features) based on Harris corner detector at multi-scale, and determine their characteristic scale. It is important to point out that the Harris corner detector is computed only once for low-resolution images while for high-resolution ones the Harris detector is employed for each scale. Features points extracted by Harris detector are described by a vector of invariant descriptors.
- *Feature matching* - Compute putative matches between image features based on the Mahalanobis distance using the invariant descriptors.
- *Geometric model estimation* – Estimate the parameter vector of a linear transformation which allows mapping the high-resolution image onto a low-resolution one.

3.1 Feature Extraction

Salient points are landmarks in an image often called interest points point of interest, and have special properties which make them stand out in comparison to its neighboring points. These features should be important dominant points of distinctive objects in images and their detection is an important task in further processing. Much of the work on two-dimensional features have concentrated on corners, that is, features formed at boundaries between two significantly dissimilar image brightness regions, where the boundary curvature is sufficiently high. However, many other types of localized structure arise, for example, ‘‘T’’,‘‘X’’ and ‘‘Y’’ junctions, and these multi-region junctions should also be considered. According to Deriche [2], the corner can be defined as image points belonging to a contour where the contour presents a local maximum curvature or as the intersection of two o more contours.

In many computer vision tasks such as: image

registration, image matching, object recognition and motion analysis, accurate corner detection is essential.

The notion of interest point was introduced for the first time by Moravec [9]. His detector is based on the autocorrelation function of the signal. It measures the gray value differences between a window and a window shifted in the four directions parallel to the rows and columns. An interest point is detected if the minimum of these four directions is superior to a threshold.

Harris and Stephens [6] defined a similar detector to that of Moravec, but whose several defects were corrected. For instance, the Moravec response is anisotropic, i.e. the local autocorrelation is only calculated in four directions, the response is noisy, as the profile of the "window" used for finding the autocorrelation is square and binary. On other hand, the Harris detector as it is always called, considers all possible directions to compute the intensity variation. The local auto-correlation function measures the local changes of the signal with patches shifted by a small amount in different directions. Besides, the "cornerness" is calculated using only the first derivatives of the intensity. A circular Gaussian smoothing window overcomes the problem caused by the rectangular binary window. This corner detector is quite robust against noise.

Comparisons between several algorithms, Schmid Bensebaa et al. [13] have shown that the Harris corner detector reaches the best repeatability rate for moderate changes of the imaging conditions. Moreover, the improved Harris version has proven invariance against image noise.

In general, there are a lot of corner detection techniques in the literature, each one with its advantages and drawbacks. However, due to its advantages (strong invariance to: rotation, illumination variation and image noise), the Harris corner detector was chosen in this work as method interest point extraction. Besides, the Harris multi-scale version proposed by Dufournaud et al. [3] is also used in this work.

In the next sections we present the Harris corner detector standard and its improved version and multiscale version.

3.1.1 Harris detector

As mentioned above Harris corner detector improved the approach of Moravec and became a popular interest point detector due to its strong invariance to: rotation, illumination variation and image noise. This detector uses the auto-correlation function to determine locations where the signal changes in two dimensions. Thus to compute a matrix related to the autocorrelation function of the image, the algorithm begins by computing the gradient of the image I in horizontal and vertical direction i.e. we compute the gradient I_x and I_y using a discrete one-dimensional mask:

$$\begin{aligned} I_x &= I * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \\ I_y &= I * \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \end{aligned} \quad (1)$$

Then, for each pixel of the image, we have the following matrix:

$$\mathbf{M} = \begin{bmatrix} A & C \\ C & B \end{bmatrix} \quad (2)$$

where

$$\begin{aligned} A &= (I_x)^2 * G \\ B &= (I_y)^2 * G \\ A &= (I_x I_y) * G \end{aligned} \quad (3)$$

In the Equation (3) G is a Gaussian used to weight the derivatives summed over the window. This Gaussian G is described as follow:

$$G(\sigma) = e^{\frac{-(x^2+y^2)}{2\sigma^2}} \quad (4)$$

Hence the matrix M can be written as:

$$\mathbf{M} = G(\sigma) * \begin{bmatrix} (I_x^2) & (I_x I_y) \\ (I_x I_y) & (I_y^2) \end{bmatrix} \quad (5)$$

The eigenvalues λ_1 e λ_2 of matrix **M** correspond to the principal curvatures of the autocorrelation function. Thus, if both eigenvalues are large, a corner is found, and if one is large and the other small, this signals an edge. The isotropic nature of the Harris detector makes it fully rotation-invariant.

To avoid explicit eigenvalues decomposition, Harris and Stephens devise a measure using the determinant and trace of the matrix **M**. Thus, the so-called corner response function ("cornerness") **R** used for corner detection is based on the determinant and the trace of matrix **M** given by:

$$\mathbf{R} = \det(\mathbf{M}) - k.\mathrm{trace}^2(\mathbf{M}) \quad (6)$$

where

$$\det(\mathbf{M}) = \lambda_1.\lambda_2 = AB - C^2 \quad (7)$$

and

$$\mathrm{trace}(\mathbf{M}) = \lambda_1 + \lambda_2 = A + B \quad (8)$$

and where the factor *k* determines the maximum ratio of eigenvalues for which **R** is positive. We use *k*=0.04 as suggested by Harris. The pixel positions of the detected points are found at local maxima of **R** above the given threshold *T* (*T* > 0).

3.1.2 Improved Harris detector

It is known that the calculation of the derivative is a complex problem, even some small noise can greatly modify the result. Therefore, for all derivation calculation it is therefore necessary to do a smoothing.

So, according to Schmid at al [13] the improved version of Harris, derivatives are computed more precisely by replacing the discrete one-dimensional mask with derivatives of Gaussian. Mathematically, this is equal to filtering the image with a Gaussian filter before calculating the gradient. This improved detector resists therefore to the noise better than the original Harris detector. Moreover, it is important to note that this detector have a better repeatability rate of interest points, in the presence of relative rotation [13]. Equation (9) and (10) represent respectively the autocorrelation matrix and the cornerness measure of the improved Harris detector.

$$\mathbf{M}_{\mathbf{imp}} = G(\tilde{\sigma}) * \begin{bmatrix} I_x^2(\sigma) & I_x(\sigma)I_y(\sigma) \\ I_x(\sigma)I_y(\sigma) & I_x^2(\sigma) \end{bmatrix} \quad (9)$$

where $G(\tilde{\sigma})$ represents the Gaussian smoothing. Remark that smoothing factor and the weighting factor are not necessarily equal.

$$\mathbf{R}_{\mathbf{imp}} = \det(\mathbf{M}_{\mathbf{im}}\mathbf{p}) - \alpha \ \text{trace}^2 \ (\mathbf{M}_{\mathbf{imp}}) \quad (10)$$

3.1.3 Multiscale Harris detector

The Harris detector is the most robust against, rotation invariance, noise, and illumination conditions, but fails in the presence of scale changes between images. In order to deal with such a transformation Dufournaud et al. [3] proposed the scale adapted Harris operator. The points are detected at the local maxima of the Harris function applied at several scales.

Thus, the interest point detector at scale *sc* is defined by:

$$\mathbf{M}_{\mathbf{s}} = sc^2 G(s\tilde{\sigma}) * \begin{bmatrix} I_x^2(s\sigma) & I_xI_y(s\sigma) \\ I_xI_y(s\sigma) & I_y^2(s\sigma) \end{bmatrix} \quad (11)$$

and the cornerness measure is the following:

$$\mathbf{R}_{\mathbf{s}} = sc^4\left(\det(\mathbf{M}_{\mathbf{s}}) - \alpha \ \text{trace}^2 \ (\mathbf{M}_{\mathbf{s}})\right) \quad (12)$$

Fig. 4 and Fig. 5 show the interest points detected using the multiscale Harris detector. In the image frame (Fig.5), the number of points detected decreases when the scale *s* increases.

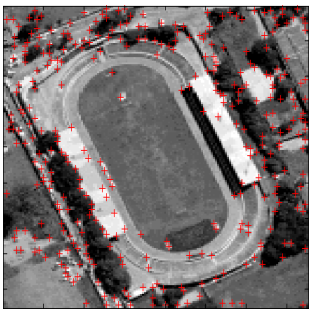

Fig.4: Georeferenced image with corners detected.

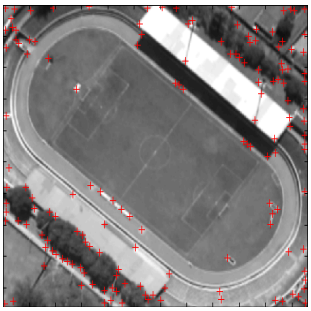

Fig.5: Frame image with corners Detected (scale 2.4).

3.2 Feature Matching

Image matching is an essential aspect of many approaches to problems in computer vision, including object recognition stereo matching, and motion tracking. In fact, several algorithms exist in the literature for the automatic point matching, even so in spite of researchers efforts, the problem is extremely complex and there isn't yet an automatic solution that gives good results in most of the cases. For this reason, many researchers are still working on this topic.

In general, techniques to find candidate matches in a pair of images mainly belong to two classes [8]:

- *The feature matching approaches*: these algorithms extract salient primitives from the images, such as edge segments or contours, and match them.
- *The iconic approaches*: these ones use directly the signal information for matching points. These approaches are based on correlation methods and differential invariants methods. The first iconic method is one of most popular techniques and has the main advantage to give very good results. However, this method is very time consuming [5]. The second iconic method is one of more recent techniques. This technique work on gray level images and consists in

characterizing points by using differential invariants of the signal.
Thus, a prominent approach to image matching has consisted of identifying "interest points" in the images, finding photometric descriptors of the regions surrounding these points and then matching these descriptors across images [1].

3.2.1 Differential invariants based matching methods

A set of derivatives in a point of the visual space, calculated until a certain order, allows characterizing the structure of the signal in the neighborhood of this point. Thus, in order to characterize the local signal in a rotationally invariant way, Koenderink and van Doorn [7] and Romeny et al. [11] computed a set of differential invariants from the " local jet". In fact, for each interest point detected, one associated a vector of a certain number of components, which are values invariant to a certain number of transformations, such as: illumination variation or rotation. In our methodology, we have limited the set of invariants to the second order. Adopting the notation of Schmid and Mohr [12], the differential invariant vector $\vec{v}$ is given by the Equations (12) and (13).
Equation (12) shows the elements of the differential invariants $\vec{v}$ in tensorial notation – the so-called Einstein summation convention while Equation (13) shows the differential invariants in Cartesian notation:

$$\vec{V} = \begin{bmatrix} L \\ L_i L_i \\ L_i L_{ij} L_j \\ L_{ii} \\ L_{i\,j} L_{ji} \end{bmatrix} \quad (12)$$

$$\vec{V} = \begin{bmatrix} L \\ L_x L_x + L_y L_y \\ L_{xx} L_x L_x + 2L_{xy} L_x L_y + L_{yy} L_y L_y \\ L_{xx} L_{yy} \\ L_{xx} L_{xx} + 2L_{xy} L_{xy} + L_{yy} L_{yy} \end{bmatrix} \quad (13)$$

3.2.2 Similarity Measure

To compare features of different interest points a distance or similarity measure is needed. The most often used measures in literature are the Euclidean and Mahalanobis distance.
The distance we used is the Mahalanobis distance given by:

$$d_M(\vec{v}_1, \vec{v}_2) = \sqrt{(\vec{v}_1 - v_2)^T \Lambda^{-1} (\vec{v}_1 - \vec{v}_2)} \quad (14)$$

where $\vec{v}_1$ and $\vec{v}_2$ v2 are two descriptors and Λ is the covariance matrix of $\vec{V}$.
The covariance matrix Λ is symmetric and positive definite then its inverse can be decomposed in the following way:

$$\Lambda^{-1} = P^T D P = P^T \sqrt{D}\sqrt{D} P \quad (15)$$

where P is an orthogonal and D diagonal matrix. Then, the Mahanalobis distance d_M can be rewritten as:

$$d_M(\vec{v}_1, \vec{v}_2)^2 = \left\| \sqrt{D} P \vec{v}_2 - \sqrt{D} P \vec{v}_1 \right\|^2 \quad (16)$$

Therefore calculate the Mahanalobis distance between two vectors is equivalent to transform these two vectors multiplying them by the matrix $\sqrt{D}P$ and then calculate the Euclidean distance between these vectors.

3.3 Geometric model estimation

3.3.1 Geometric Transformation

After the feature correspondence has been established (second step of the methodology), the task consists of choosing the type of the mapping function and its parameter estimation.
Thus the geometric model used is based on similarity transformation. A basic image similarity-based method consists of a transformation model, which is applied to test image coordinates to locate their corresponding coordinates in the georeferenced image.
In this work, the idea behind matching methods is to search for the best matching while permitting the rotation, translation and scaling (to be called alignment by similarity transformation)
Let (u,v) be the point coordinates of georeferenced image and let (x,y) be the point coordinates of frame image.
Then a point (x,y) present in frame image is related to point (u,v) of the georeferenced image by linear geometric transformation as follows:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos\theta & -sin\theta \\ sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} s & 0 \\ 0 & s \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (17)$$

where s is an arbitrary scaling factor, θ is an arbitrary rotation and (t_x,t_y) is an arbitrary shift.

The Equation (17) becomes

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (18)$$

with $a = s.\cos\theta$ and $b = s.\sin\theta$

The four unknown parameters of our model are: a, b, t_x, t_y..

Then, we have a linear system as:

$$\mathbf{A}x = b \quad (19)$$

The similarity transformation, Equation (18) which have four unknown parameters is estimated by solving the system of Equation (19). Note that for each point the transformation gives two equations, one for each coordinate.

Thus, the best way to solve the system of the Equation (19) is to perform singular value decomposition of the matrix A.

The least-squares solution for the parameters **x** can be determined by solving the corresponding normal equations,

$$x = \left[\mathbf{A}^{\mathrm{T}}\mathbf{A}\right]^{-1}\mathbf{A}^{\mathrm{T}}b \quad (20)$$

which minimizes the sum of the squares of the distances from the model locations to the corresponding matched points locations.

3.3.2. RANSAC Algorithm

The key idea of this algorithm is to match images at different scales. Since there is a strong relationship between scale and resolution, the method proposed in this work allows automatically matching points between two images with different resolutions (the high-resolution image and low-resolution one. During the matching process the scale factor related with the high-resolution image is chosen to increase monotonically and this process attempts to find which one of these images best matches a region in the low-resolution image. The similarity transformation shown in Equation (18) is reliable to find the parameters, which allows finding the optimal coordinates of estimated position of the UAV. However, as we need to select the best correspondents points between two images at different scales, we need a robust algorithm for this task. In this sense the RANSAC algorithm proposed by Fischler and Bolles [4] takes as input the potential one-to-one point assignments, computes the best transformation between the two images, and splits the point assignments into two sets: inliers and outliers [3].

The RANSAC algorithm (RANdom SAmple Consensus) is most popular approach for this robust estimation problem algorithm, even though several variants have been developed. This algorithm facilitates to distinguish inliers matching points and outlier ones. Thus, beginning from a subset of points the parameters of transformation model are estimated and the number of inliers and outliers are computed. Note the smaller the initial subset, the smaller the probability to detect outliers. This process, repeats until itself reached a certain degree of satisfaction: for instance a probability (usually 95%) to have chosen a subset of inliers after N trials.

Thus, the RANSAC algorithm is based on the following steps:

1. Randomly select a sample of s data points from S and instantiate the model from this subset;
2. Estimate the parameters a, b, t_x, t_y of the model (Equation 18);
3. Determine the set of data points S_i which are within a distance threshold d of the model. The set S_i is the consensus set of samples and defines the number of inliers of S_i and save it and save the estimated parameters a, b, t_x, t_y;
4. If the subset of S_i is greater than some threshold d, re-estimate the model using all the points in S_i and terminate;
5. If the size of S_i is less than d, select a new subset and repeat the above.
6. After N trials the largest consensus set S_i is selected, and the model is re-estimated using all the points in the subset S_i.

There are three important parameters to define in the RANSAC algorithm: the number of s samples, the threshold value d and the number of trials N.

- The first one (number of samples s) must be smaller since it allows to minimize the detection of outliers;
- The second one (the threshold d) must be chosen according of measures characteristic. In the RANSAC algorithm, only the points whose Eulidean distance is lower to threshold d are kept. When we process data with noise, it is preferable to use Mahalanobis distance than Euclidean distance;
- The third parameter is the number of the trials N.

The ideal number of trials is to consider all possible combinations. However this can be very expensive in calculation time. The number of N must be taken sufficiently great to have p probability, often equal to 0.99 that allows to have one of the trials N which results to have none oulier. This number depends therefore greatly of the proportion of inliers points in the data. The next Equation gives the number of the trials N related to probability p and the proportion ε of the outliers contained in the data. …

$$N = \log(1-p)/\log((1-(1-\varepsilon)^s) \quad (21)$$

The five steps of the RANSAC algorithm are utilized for each scale. Then, for each scale we save the greatest number of inliers and the estimated parameters of the model. Finally we choose which scale has the greatest number of inliers and we use the Equation (17) to map the high-resolution image onto a low-resolution image, which result on UAV position estimation.

.

4. Results

In our experiments we have used the following scale factors: 1.4, 1.8, 2.4, 2.8, 3.4, 4.4.

We have obtained the highest value of inliers using the scale factor 2.4.

By using the similarity transformation (Equation 18) we have obtained the estimated position in the georeferenced image (Fig. 8) based on the coordinates of the frame image in Fig. 7.

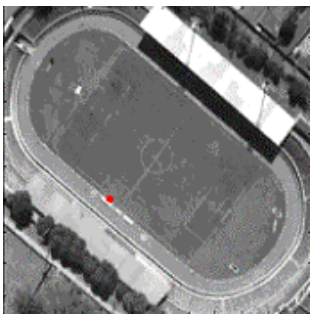

Fig.7: Frame image after mapping.

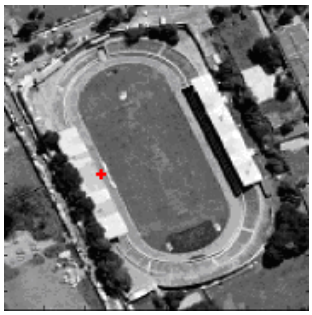

Fig.8: Georeferenced image after mapping

5. Conclusion

An algorithm for detection interest points, matching them and the use a geometric model to mapping two images was presented in this paper. Our primary motivation behind the development of this algorithm was its use in reliable and efficient localization estimation for autonomous aerial navigation. Due to the difference in terms of resolution between frame images and georeferenced images, the methodology proposed was to consider the matching process in different scales. Experimental results shown that the parameters estimated allowed a reasonable performance of the UAV position estimation.

In the work described in this paper, we have assumed that the aerial images are taken from a nadir view. However, future works will address to investigate approaches using images taken from oblique view.

References

[1] Benhimane, S.; Malis, S. *Matching images at different resolutions using intrinsics-free measures.* 14eme Congres Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Toulouse, France, January 2004.

[2] Deriche, R.; Giraudon, G. *A Computational Approach for Corner and Vertex Detection* International Journal of Computer Vision. v. 10, p. 101-124, 1993.

[3] Dufournaud, Y.; Schmid, C.; Horaud, R. *Matching Image with Different Resolutions.* Computer Vision and Pattern Recognition, v. 1,612-618, 2000.

[4] Fischer, R.; Bolles, R. Random sample consensus: *A paragdigm for model fitting with application to image analysis and automated cartography*. Proceeding**…**Graphics and Image Processing, 24(6): 381 – 395, 1981.

[5] Gouet, V.; Montesinos, P.; Pelé, D. *Stereo matching of color images using differential invariants. In* Proceedings of the International Conference on Image Processing, Chicago, USA, *1998.*

[6] Harris, C.; Stephens, M. *A combined corner and edge detector*. In: Alvey Vision Conference, 1988.

[7] Koenderink, J.; van Doorn, A. J. *Representation of local geometry in the visual system.* Biological Cybernetics, v. 55, p. 367-375, 1987.

[8] Montesinosa,P.; Gouet, V.; Deriche, R.; Pele, D. *Matching color uncalibrated images using differential invariants.* Image and Vision Computing. v. 18 p. 659–671, 2000.

[9] Moravec, H. P. *Toward automatic visual obstacle avoidance.* In: Proc. Int. Joint Conf. Artificial Intelligence, p. 84–584, 1977.

[10] Nilsson, J. Visual *Landmark Selection and Recognition for Autonomous Unmanned Aerial Vehicle Navigation.* Master's Degree Project Stockholm, Sweden 2005.

[11] Romeny ter Haar, B. M.; Florack, L. M. J.; Salden, A. H.; Viergever M. A. *Higher order differential structure of images*. Image and Vision Computing. v. 12, p. 317-325, July/August 1994.

[12] Schmid, C.; Mohr, R. *Matching by local invariants*. Technical Report 2644, INRIA, 1995.

[13] Schmid, C.; Mohr, R.; Bauckhage, C. *Evaluation of Interest Points Detectors*. In: International Journal of Computer Vision, 37(2), p. 151-172, 2000.