UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Visión Electrónica
Más que un estado sólido
https://doi.org/10.14483/issn.2248-4728

VISIÓN ELECTRONICA

RESEARCH VISION

Recognition of objects with feature matching and RANSAC algorithm

Reconocimiento de objetos implementando características puntuales y el algoritmo RANSAC

Edgar Steven Correa-Pinzón[1]

INFORMACIÓN DEL ARTICULO

Historia del articulo
Enviado: 23/09/2018
Recibido: 24/09/2018
Aceptado: 12/11/2018

Keywords:
Image-processing
Image-Tracking
Augmented-reality
Homography.

Palabras clave:
procesamiento
seguimiento,
imágenes
realidad aumentada
homografía.

ABSTRACT:

This paper shows some results of research works: first, the implementation of artificial vision techniques for the treatment of images, such as: filtering, edge detection, morphological operations, location and recognition, and secondly the implementation of the point of interest invariant detector to scale and rotation. descriptor together with the *Ransac* probabilistic method to derive an object matching methodology. There are two cases of application, one of these in the domain of augmented reality. The object identification methodology that is implemented has excellent results, even with cases of occlusion; however, for MATLAB implementation it is desirable to increase the processing speed almost in real time and implementation in practical cases.

RESUMEN

Este artículo muestra algunos resultados de trabajos de investigación: en primer lugar, la implementación de técnicas de visión artificial para el tratamiento de imágenes, tales como: filtrado, detección de bordes, operaciones morfológicas, localización y reconocimiento; en segundo lugar, la implementación del detector y descriptor de punto de interés invariante a escala y rotación junto con el método probabilístico *Ransac* para derivar una metodología de coincidencia de objetos. Se presentan dos casos de aplicación, uno de estos en el dominio de realidad aumentada. La metodología de identificación de objetos que se implementa presenta excelentes resultados, incluso con casos de oclusión; sin embargo, para la implementación en MATLAB es deseable aumentar la velocidad de procesamiento casi en tiempo real y la implementación en casos prácticos

1BSc. In Automation Engineering, Universidad de la Salle, Colombia. E-mail: ecorrea00@unisalle.edu.co ORCID: https://orcid.org/0000-0002-1477-6825
Cite this article as: E. S. Correa-Pinzón "Recognition of objects with feature matching and RANSAC algorithm", *Visión electrónica*, vol. 2, no. 1, Special edition, January-June 2019

1. Introducción

Computer vision is the transformation of data from a still or video camera into either a decision or a new representation [1] data may be multiple photographs, data from different sensors, from different times, or from different viewpoints, all of these transformations are done for achieving some particular goal, recently there have been developing with greater impact applications in analytical photogrammetry such as cartography, military intelligence, city planning and other applications that include interactive image synthesis in the augmented reality domain [2] as well mobile robot navigation, human-computer interaction and medical image analysis. Due to the potential in this field, it has become an important topic of applied research, where it is possible to develop applications and even more important to show the way in which it's implemented.

For the study case assume that we have a region in the space, and a supposition that some interest object might be present in this region with a particular position and orientation, the solution has been obtained using image processing methods specially the scale and rotation invariant interest point detector and descriptor together with the probabilistic method Ransac. Two cases of application are presented, the first consist in a methodology for object detection and the second it's a sudoku solver in the augmented reality domain.

The article is structured like follow. First, the methods: Feature matching the implementation of RANSAC and Homography; then, the results: Recognition of objects, and Sudoku solver with computer vision; and, finally, the conclusions.

2. Methods

There are two ways to generate image feature matching, one of this is intensity base method, which consists in compare intensity patterns in images via correlation metrics registering entire image or sub-images, and the other is feature-based method, which consists in find correspondence between image features such as interest points and descriptors, establishing point by point correspondence between the reference and target images, the latter has been using in the work methodology.

The search for discrete image correspondences can be divided into three main steps [3] First, 'interest points' are selected at distinctive locations in the image, such as corners, blobs, and T-junctions. The most valuable property of an interest point detector is its repeatability, i.e. whether it reliably finds the same interest points under different viewing conditions. Next, the neighborhood of every interest point is represented by a feature vector. This descriptor must be distinctive and, at the same time, robust to noise, detection errors, and geometric and photometric deformations. Finally, the descriptor vectors are matched between different images. The matching is often based on a distance between the vectors [3]. There are a lot of descriptors, and there are some researches about the performance of the most famous like SIFT, SURF, DAISY, ORB, BRIEF, [4 - 5].

The dimension of the descriptor has a direct impact on the time this takes, and a lower number of dimensions is therefore desirable, but also sacrificing performance. One of the first objectives is to find the correct correspondences between the images as illustrated in Figure 1. In this case, in addition to generating correspondences with the object of interest in the scene, some correspondences with other objects are also generated, however, it is correct because in this case they are identical characteristics, in specific the number twelve is repeated in the scene.

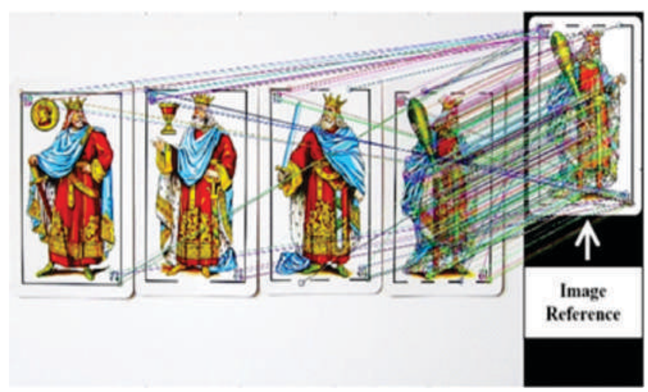

Figure 1. Correct correspondences between images. Source: own.

To achieve the feature matching as illustrated in Figure 1, it is necessary to work with the information of all descriptors that are shown in Figure 2 where there are correspondences that aren't correct, this is because there are fewer descriptors in the reference image in relation to the scene, since it is assumed that the scene in addition to containing the object of interest has more objects.

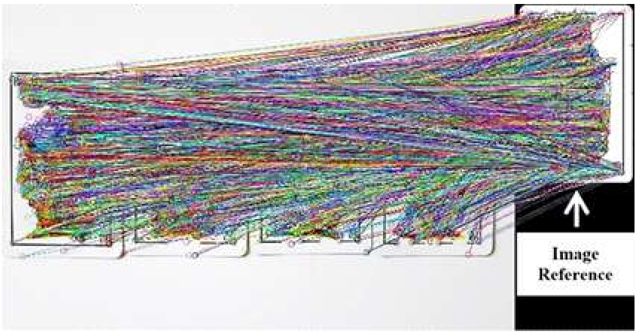

Figure 2. Correspondences between images. Source: own.

2.1. Feature matching

To find this feature matching between the images, as a first step is implemented matching by force strategy like as presented in part (a) of Figure 3, that consist in working in matrix way with all descriptors due to descriptors are in a meta-space, where it's desirable to determine what descriptor in the scene correspond with which descriptor of the reference image, for this is done arccosine(A(i)*B) where A contains all descriptors in the scene and B contains all descriptors in the reference image but transposed, as a result a fitness vector is obtained "match", and obviously the value with the best attitude is the position of the descriptor of interest, the implementation of "distRatio" through an "IF" is a guarantee of having contracted an outstanding descriptor.

After force strategy it's important the implementation of RANSAC, the probabilistic method explains the data using models that are global. These models try to explain a large data collection with a small number of parameters. For example, we could take a couple of images and try to fit a parametric set of motion vectors that explain how the pixels move from one to the other, which first generates hypotheses from subsets of observations and then evaluates these hypotheses against all the available observations in order to determine which hypothesis is the best representative of the entire population (the closest from the ideal solution) and keeps working until it reaches a confidence threshold [6].

Modifications about RANSAC have proposed, with experimental results, both from synthetic and real data have testified the efficiency [7] the authors call this method as outlier elimination based RANSAC consists in after obtaining a plausible estimate, the obvious outliers are eliminated from the set of correspondences with the objective of improving the index of values in the set of remaining correspondences, which will accelerate the progress of the sample.

In this case the implementation of RANSAC is presented in part (b) of Figure 3, processes the coordinates of the feature matching, obtained in matching by force strategy. Four coordinates (x, y) of MatchTable are randomly selected (I1 for Scene image and I2 for Reference image), and generate the homography matrix, now the fitness of the homography is determined, this through the euclidean distance, to determine the Euclidean distance, it's calculated the difference of the square between the coordinate of the scene image and the coordinate of the reference image transformed with the homography H, with the parameter "DistF" determine the amount of inlier desired, this is the way to measure the fitness of calculated homography for each iteration "the best homography is the one that contains the most inliers". It is important to note that the "thInlr" parameter generates the discarding of homographies that do not have the percentage of inliers that are defined with "thInlrRatio".

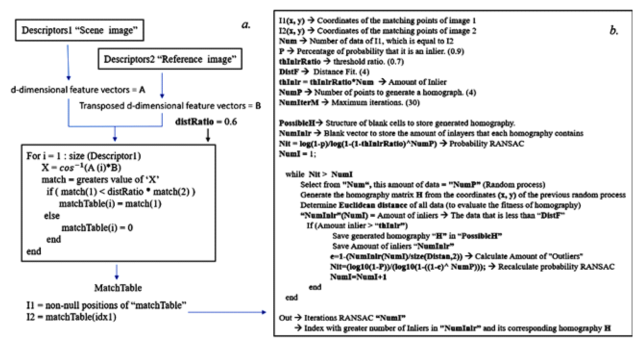

Figure 3. (a) Matching by force strategy, (b) RANSAC implementation. Source: own.

2.2. Homography

Linear transformations models are used to relate the target image space to the reference image space. Included are translation, rotation, scaling, general affine and projective transforms. Linear transformations are global in nature thus; they cannot model local geometric differences between images.

Euclidean transformation involves a rotation and plane translation, another similarity involves a scaled rotation and translation that need 2 points correspondences to be estimated shown in (1). In six degrees of freedom Parallel lines are preserved that need 3 points correspondences to be estimated shown in (2), and in homography or eight degrees of freedom straight lines are preserved that needs 4 correspondences to be estimated shown in (3).

$$\begin{vmatrix} X_T \\ Y_T \\ 1 \end{vmatrix} = \begin{vmatrix} S.\cos\Theta & -S.\sin\Theta & t_x \\ S.\sin\Theta & S.\cos\Theta & t_y \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} X_R \\ Y_R \\ 1 \end{vmatrix} \tag{1}$$

$$\begin{vmatrix} X_T \\ Y_T \\ 1 \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} X_R \\ Y_R \\ 1 \end{vmatrix} \tag{2}$$

$$\begin{vmatrix} X_T \\ Y_T \\ Y_T \end{vmatrix} = \begin{vmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{21} & h_{23} & h_{33} \end{vmatrix} \begin{vmatrix} X_R \\ Y_R \\ 1 \end{vmatrix} \tag{3}$$

Each point to point correspondences give rise to two independent equations, shown in (4)

$$X_T = H.X_R \tag{4}$$

Matrix h can be written as a vector h, shown in (5)

$$\mathrm{h} = |h_{11}\ h_{12}\ h_{13}\ h_{21}\ h_{22}\ h_{23}\ h_{31}\ h_{32}\ h_{33}| \tag{5}$$

$X_T = HX_R$ Can be expressed as a linear system Ah=0

$$\mathrm{A} = \begin{vmatrix} X_{R1} & Y_{R1} & 1 & 0 & 0 & 0 & -X_{R1}.X_{T1} & -Y_{R1}.X_{T1} & -X_{T1} \\ 0 & 0 & 0 & X_{T1} & Y_{T1} & 1 & -X_{R1}.Y_{T1} & -Y_{R1}.Y_{T1} & -Y_{T1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{Rn} & Y_{Rn} & 1 & 0 & 0 & 1 & -X_{Rn}.X_{Tn} & -Y_{Rn}.X_{Tn} & -X_{Tn} \\ 0 & 0 & 0 & X_{Tn} & X_{Tn} & 1 & -X_{Rn}.Y_{Tn} & -Y_{Rn}.Y_{Tn} & -Y_{Tn} \end{vmatrix} \tag{6}$$

Inhomogeneous solution, one of the nine matrix elements is fixed to unity value (h33). A new equation of the form A'.h'=b is formed, shown in (7)

$$\begin{vmatrix} X_{T1} \\ Y_{T1} \\ \vdots \\ \vdots \\ X_{Tn} \\ Y_{Tn} \end{vmatrix} = \begin{vmatrix} X_{R1} & Y_{R1} & 1 & 0 & 0 & 0 & -X_{R1}.X_{T1} & -Y_{R1}.X_{T1} \\ 0 & 0 & 0 & X_{T1} & Y_{T1} & 1 & -X_{R1}.Y_{T1} & -Y_{R1}.Y_{T1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{Rn} & Y_{Rn} & 1 & 0 & 0 & 1 & -X_{Rn}.X_{Tn} & -Y_{Rn}.X_{Tn} \\ 0 & 0 & 0 & X_{Tn} & X_{Tn} & 1 & -X_{Rn}.Y_{Tn} & -Y_{Rn}.Y_{Tn} \end{vmatrix} \begin{vmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{vmatrix} \tag{7}$$

As shown in (8), it solves using a Gaussian elimination in the case of a minimal solution or using a pseudoinverse method in case of an overdetermined system.

$$h = (A'^T.A')^{-1}A'^T.b' \tag{8}$$

With the implementation of the RANSAC algorithm, it is desirable to obtain a homography matrix that describes the exact position in size and orientation, scaling, rotation, translation and so on, with any perspective. In this case, RANSAC is used for tracking: only project the reference frame on the current image.

Methodology application for sudoku solver in the augmented reality domain is presented in Figure 4, first, it's seeking to recognize the sudoku with the methodology of object recognition presented.

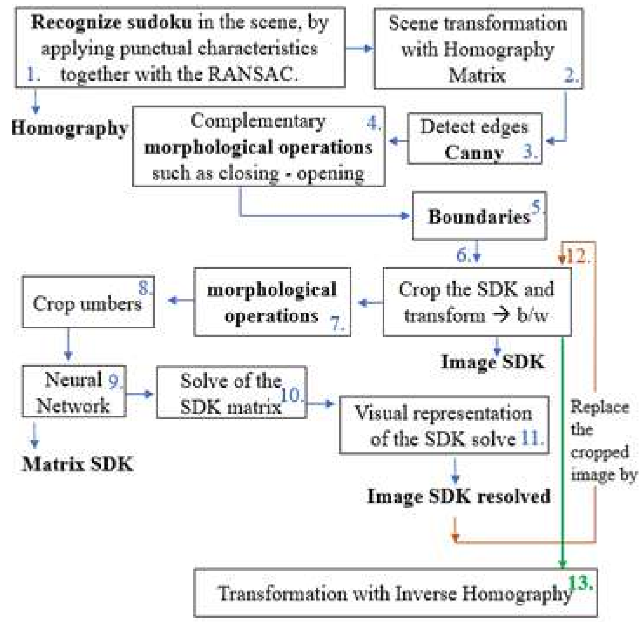

Figure 4. Proposed methodology solving Sudoku. Source: own.

After, making use of the matrix of homography and image processing techniques, along with tools such as neural networks to identify numbers and a sudoku solver, is obtained an integration to solve the sudoku with computer vision.

3. Results

Recognition of objects, once the descriptors of each point characteristic, filtered by force strategy and implemented the presented algorithm, are transformed in the four corners of the reference image with the homographic matrix to obtain the results shown in Figure 5-8, is where the identification of the interest object in the proposed scenes is observed, it's convenient to highlight the problem of occlusion, in which the methodology developed behaves in an efficient manner.

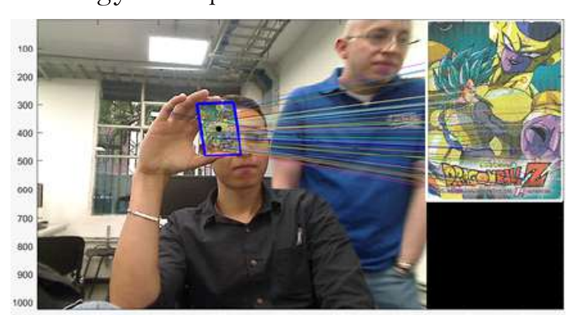

Figure 5. Recognition of an object in a space. Source: own.

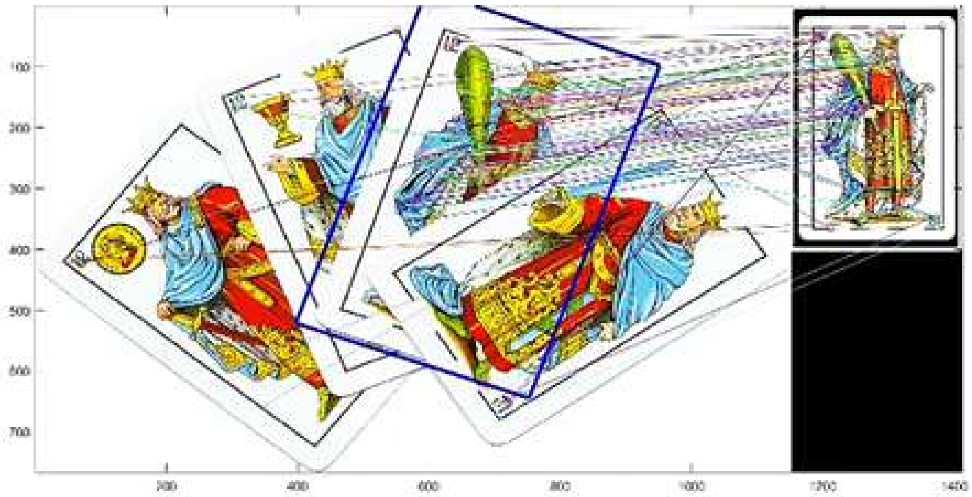

Figure 6. Recognition of an object with occlusion. Source: own.

Figure 7. Recognition of objects with occlusion. Source: own.

Figure 8. Face recognition [8].

	key points Image_Scene	key points Image_Reference	Matches
Figure 1	3853	663	139
Figure 2	3425	663	3425

Table 1. Number of descriptors and matches**.** Source: own.

Figure 9. Sudoku solver in the augmented reality domain. Source: own.

3.1 Sudoku solver with computer vision

Step one of the Figure 4, as shown in Figure 10 is found the sudoku position.

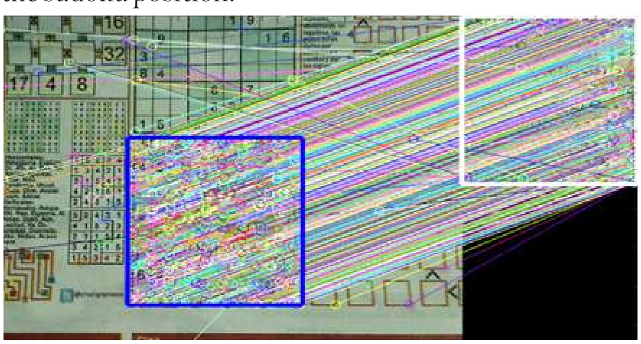

Figure 10. Identification of a sudoku position in the scene. Source: own.

Step two of the Figure 4, as shown in Figure 11 is transformed with homography, to work with the straightened sudoku.

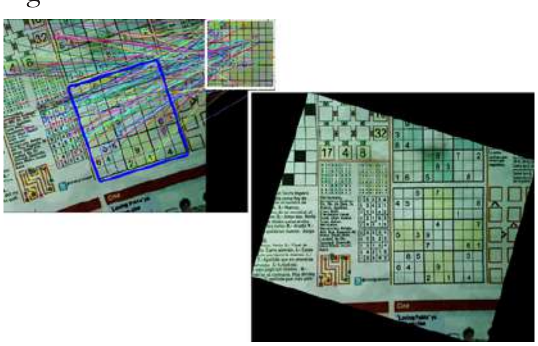

Figure 11. Image is transformed with homography. Source: own.

Step three to six of the Figure 4, as shown in Figure 11 is obtained the black and white sudoku is obtained to evaluate the numbers of each position in a neural network to represent as numbers the information of the pixels that make up the numbers.

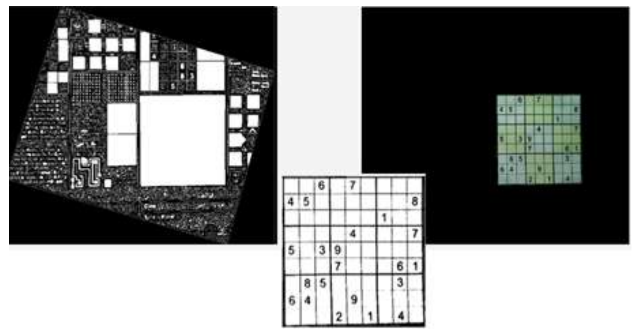

Figure 12. Sudoku recognized. Source: own.

In Figure 13 is shown the fraction of a data base created to train the neural network the step nine of the Figure 4

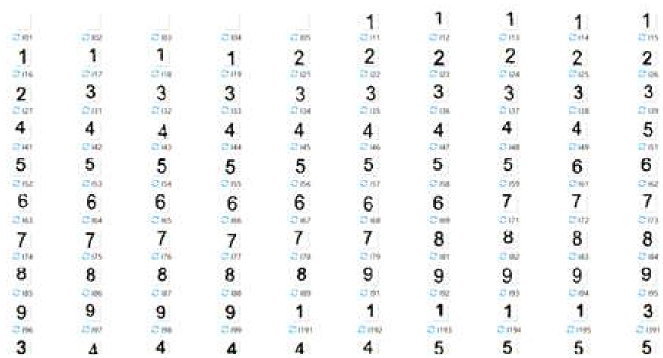

Figure 13. Data base for training the neural network. Source: own.

With whatever sudoku resolution tool it's solved, step ten of the Figure 4. To obtain a visual representation of the SDK solve like is shown in Figure 14, The pixels representing a zero are replaced with the pixels representing the resolved information of the sudoku. (are replaced the pixels of the image R channel, so that the numbers are displayed in red).

Figure 14. Visual representation of the SDK solve. Source: own.

Step twelve of the Figure 4, is shown in Figure 15

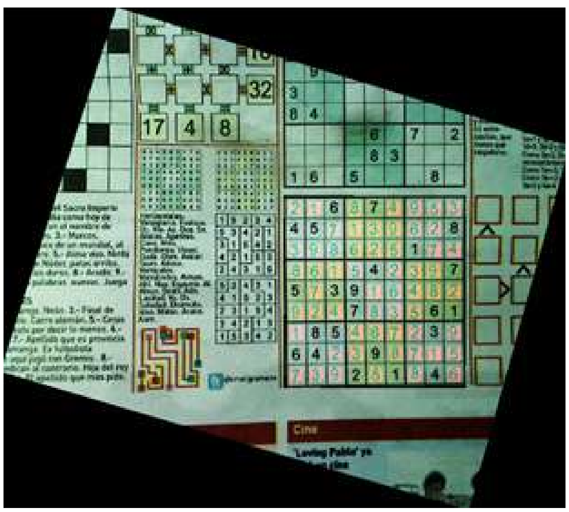

Figure 15. Replace the visual representation of the SDK solve in the scene. Source: own.

Figure 16. Transformation with Inverse Homography. Source: own.

The results of the integration of image processing techniques are shown in Figure 9, 16

4. Conclusions

The tasks carried in automated motion estimation are computationally expensive, resulting in a considerable amount of time dedicated to performing them. Thus, one should concentrate in achieving accurate estimations in a not prohibitive amount of time when planning a computer vision system. In other words, one needs to reach the balance between these goals in order to design a good system [6].

The descriptors of each specific characteristic have been obtained, the dimension of the descriptor has a direct impact on the time this takes, and a lower number of dimensions is therefore desirable, but also sacrificing performance. In [Table 1] the effectiveness of the implementation of the strategy of force is shown, reducing from 3425 matches to only 139, lightening the load at the time of implementing RANSAC.

In some cases, the standard algorithm has insufficient performance. Thus, several methods were published targeting the goal of speeding up RANSAC tasks and dealing with the trade-off between quality and time by working in three major points: alternative hypothesize and verify methods, improvements on the hypotheses generation and improvements on the hypotheses [6].

The object identification methodology that is implemented in this paper, presents excellent results as can be seen in Figure 5-8, even with occlusion cases. However, the implementation is done in MATLAB, it's desirable to increase the processing speed almost real time and implementation in practical cases, then it's projected as future work to do the implementation in python with a raspberry-pi.

Because we are such visual creatures, it is easy to be fooled into thinking that computer vision tasks are easy, how hard can it be to find, anything in an image. The human brain divides the vision signal into many channels that stream different kinds of information into your brain, the brain has an attention system that identifies in a task-dependent [9].

The artificial systems look for the similarity with the natural processing of images, that is why the present work of research seeks to generate an approach to the resolution of tasks in the field of computer vision in a functional and efficient way, with the aim of developing applications increasingly complex.

Acknowledgments

Mainly to the professor Cesar Hernan Rodríguez Garavito, Ph.D., for everything learned in the chair of computer vision & flexible computing, and his contribution in a professional training.

References

[1] B. Gary, and A. Kaehler. "Learning OpenCV", Computer vision with the OpenCV library, O'Reilly Media, Inc.", 2008.

[2] D. Forsyth and J. Ponce, "Computer Vision: A Modern Approach", Englewood Cliffs, NJ:Prentice-Hall, 2003.

[3] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features", European conference on computer vision. Springer, Berlin, Heidelberg, 2006. https://doi.org/10.1007/11744023_32

[4] Mikolajczyk, Krystian, and Cordelia Schmid. "A performance evaluation of local descriptors." *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005. https://doi.org/10.1109/TPAMI.2005.188

[5] Zakharov, A. A., A. Yu Tuzhilkin, and A. L. Zhiznyakov. "Finding correspondences between images using descriptors and graphs", *Procedia engineering*, vol. 129, pp. 391-396, 2015. https://doi.org/10.1016/j.proeng.2015.12.131

[6] S. Gomes-Nieto, and B. M. Carvalho, "Generalized preemptive RANSAC making preemptive RANSAC feasible even in low resources devices", International Conference on Computer Vision Theory and Applications (VISAPP), 2014.

[7] S. Yang, and B. Li. "Outliers elimination based RANSAC for fundamental matrix Estimation", International Conference on Virtual Reality and Visualization, 2013. https://doi.org/10.1109/ICVRV.2013.63

[8] Revista Semana, Orador por excelencia. Publicado el 17 ago. 2014.

[9] Z. Bo. "Computer vision vs. human vision." Cognitive Informatics (ICCI), 9th IEEE International Conference, 2010. https://doi.org/10.1109/COGINF.2010.5599750