VISUAL OBJECT DETECTION BY COLOR, SHAPE AND DIMENSION

DETECCIÓN VISUAL DE OBJETOS POR COLOR, FORMA Y DIMENSIÓN

Alejandro Israel Barranco Gutiérrez
Cátedras CONACyT – Tecnológico Nacional de México / IT de Celaya, México
israel.barranco@itcelaya.edu.mx

Saúl Martínez Díaz
Tecnológico Nacional de México / IT de la Paz, México
saulmd@itlp.edu.mx

Juan Prado Olivarez
Tecnológico Nacional de México / IT de Celaya, México
juan.prado@itcelaya.edu.mx

Recepción: 26/octubre/2020 **Aceptación:** 27/noviembre/2020

Abstract

3D neural object detection by color, shape and dimension (3DOD-CSD) is a novel and powerful tool that can be used to build object detection systems in environments with uncontrolled illumination. Inspired by the global structure of the human visual system, it uses a neural network in the classification stage and determines the physical dimension of the object's fea-tures using commercial digital cameras calibrated in stereo configuration. This permits the analysis of images of objects with the same form and color but different dimensions, such as scaled replicas or photographs of a photograph of a 3D object. With this method, a fixed distance from the camera to the object to be analyzed is not necessary - essential for a dynamic recognition system in changing conditions. The results show strong discrimination between desired and undesired objects. This system has many possible applications, including face identification and object selection in varying environments, utility pole detection, coin detection, and more.

Keywords: Color, dimensional features, invariant features, neural network, stereo vision, 3D object recognition.

Resumen

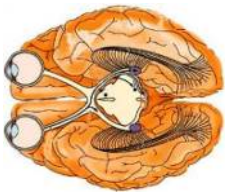
La detección de objetos 3D por color, forma y dimensión (3DOD-CSD) es una herramienta para construir sistemas de detección de objetos en entornos con iluminación incontrolada. Inspirado en la estructura global del sistema visual humano, utiliza una red neuronal en la etapa de clasificación y determina la dimensión física de las características del objeto utilizando cámaras digitales comerciales calibradas en configuración estéreo. Esto permite el análisis de imágenes de objetos con la misma forma y color, pero de diferentes dimensiones, como réplicas a escala o fotografías de una fotografía de un objeto 3D. Con este método, no es necesaria una distancia fija entre la cámara y el objeto a analizar, algo esencial para un sistema de reconocimiento dinámico en condiciones cambiantes. Los resultados muestran una fuerte discriminación entre objetos deseados y no deseados. Este sistema tiene muchas aplicaciones posibles, incluida la identificación de rostros y la selección de objetos en entornos variantes, detección de postes, detección de monedas, entre otros.

Palabras Clave: *Color, características dimensionales, características invariantes, reconocimiento de objetos 3D, red neuronal, visión estéreo.*

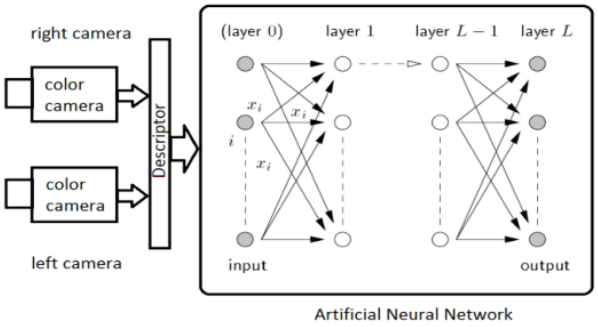
1. Introduction

In recent years automatic 3D object detection has been an area of intensive research. Several pattern recognition techniques allow objects in images to be discriminated by texture [Zhuang, 2013], [Marquez, 2016], shape [Sossa, 2006], [Wang, 2008], or colored features [Kviatkovsky, 2013]; however, the majority require controlled conditions and constant variables, such as illumination levels, distance from camera to target, or known object position on the image (image registration). These requirements limit the applicability of the techniques in uncontrolled conditions. Thus far, systems in the literature are unable to replicate human visual activities such as distinguishing be-tween a real object and its photograph, recognizing two identical 3D bodies of different scales, or recognizing images of the same object at varying distances from the camera. This paper presents an approach inspired by the global structure of the human visual system, solving these problems

by integrating dimensional features (DF), color, and shape into the recognition process and classifying objects through an Artificial Neural Network (ANN). This scheme is illustrated in figure 1, which show the analogue structure of human vision and 3DNOD-CSD.



a) Eyes connected by neurons directly to human brain



b) Cameras connected by the descriptor to Artificial Neural Network

Figure 1 Structure of human vision and 3DNOD-CSD.

3D dimensional measurement has demonstrated itself to be very effective in dealing with the limitations of 2D images, such as variations in pose and illumination, as demonstrate the works [Mustafah, 2012], [Rahman, 2009], [Xiaoming, 2009], [Amos, 2001], [Gupta, 2011], [Baek, 2010].

Shape is used as the main distinguishing factor of objects in images [Sossa, 2006], [Hu, 1962], and colour discrimination is widely used as a reliable feature in identifying objects [Yi-Xing, 2007], [Ren, 2009], [Barranco, 2013]. This proposal is based on a stereoscopic system that ensures 3D recognition from cap-tured information. The configuration is flexible enough for a dynamic recognition system in which the object is photographed at an unknown distance from the camera [Zhang, 2000]. The HSV model of color analysis is used, in which H describes the color in terms of an angle and V itemizes the illumination [Kviat-kovsky, 2013]. And respect to the object shape discrimination the Hu invariant moments are used as [Sossa, 2006]. This paper

proposes the use of (DF) to define the invariant physical dimensions of a specific object. For this system, readily available stereoscopic video cameras (3D camcorders) can be used; however, it is also possible to calibrate a stereoscopic system from two conventional cameras, using Zhang's method [Zhang, 2000], [Hartley, 2003], analyzed by Barranco [Barranco, 2009]. The document is organized as follows: Section 2 presents related work, Section 3 shows the system configuration, Section 4 explains the fundamentals of the 3DNOD-CSD algorithm, Section 5 describes experiments and results, Section 6 opens the Discussion and finally Section 7 exposes the author's conclusions.

2. Methods

3DOD-CSD has five stages requiring calibration:

- Individual camera calibration.
- Stereo calibration.
- Color segmentation.
- Shape description by Hu moments.
- Dimensional features extraction.
- ANN training.

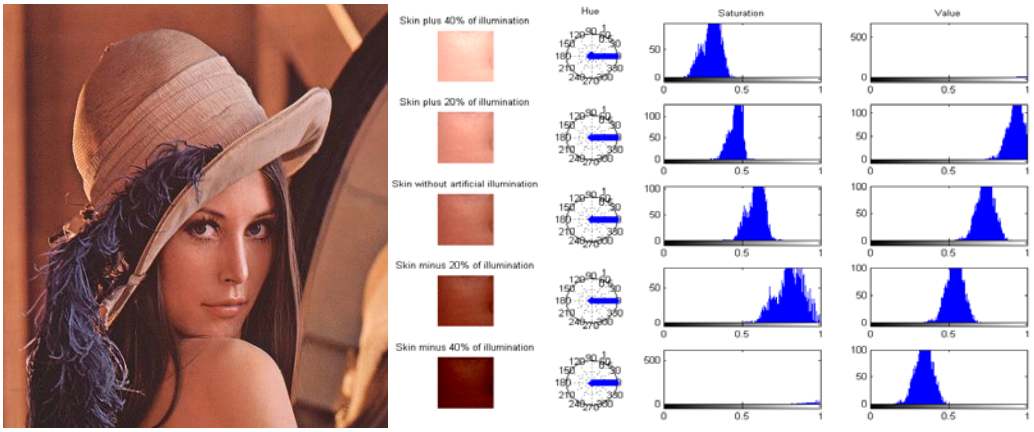
Some constants are necessary for definition according to the situation.

Stereo camera calibration

Image acquisition is performed by calibrated stereoscopic cam-eras as described in [Zhang, 2000] and [Hartley, 2003]. Using Zhang's method for each camera, the intrinsic parameters are computed. Next, the stereo cameras are calibrated - as described by Harley in [Hartley, 2003] - through isolating a reference point in left and right cameras to calculate the translation vector and rotation matrix; calculations are done with Essential and Fundamental matrices, after matching between left and right images. Next, the linear transformation between the right and left cameras' reference systems is used to triangulate 2D correspondence points in left and right images, determining their corresponding 3D point [Barranco, 2013].

Color segmentation

In [Kviatkovsky, 2013] [Ren, 2009], [Barranco, 2013] and [Sigal, 2004], color image analysis is performed with the HSV color model. [Kviatkovsky, 2013] and [Sigal, 2004] define some coefficients that depend on divisions between Hue, Saturation, and Value components, but the change in one component changes the quotient, thus when Value is used, it cannot be employed as an invariant color description. Therefore, we propose to return to the original objective of the HSV model: color perception in varying levels of illumination which mimics human color perception. Hue describes the pixel's color in an image, Saturation expresses the color quantity, and Value is directly related to physical white illumination, which may be considered to be the main descriptor of illumination and is useful to find illumination dependence. Figure 2 presents skin color analysis using the HSV model: a skin segment is artificially illuminated to observe the effect on HSV components; it can be observed that the Hue histogram is invariant to artificial white illumination changes. Examining Hue histograms for the same skin with different illuminations measured in luxes, it can be seen that they remain invariant to changes in illumination.



a) Source: Dwight Hooker *b) Source: The authors.*

Figure 2 HSV model histograms of one segment of Lena image.

Therefore, in this proposal, the central limit theorem is applied to describe the color histogram as a normal distribution N(a,σ) after computing the statistics of a large set of skin image segments [Box, 2005]. Figure 3 shows the invariance of skin color of

the Hue component of the HSV model for real changes in illumination in a typical indoor illumination. Where the DF measured as the distance between (x_0, y_0, z_0) and (x_1, y_1, z_1) can be translated by vector $t = \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix}$ and rotated respect to X, Y and Z axes by $R_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & \text{sen}(\alpha) \\ 0 & -\text{sen}(\alpha) & \cos(\alpha) \end{bmatrix}$, $R_\beta = \begin{bmatrix} \cos(\beta) & 0 & -\text{sen}(\beta) \\ 0 & 1 & 0 \\ \text{sen}(\beta) & 0 & \cos(\beta) \end{bmatrix}$ and $R_\theta = \begin{bmatrix} \cos(\theta) & -\text{sen}(\theta) & 0 \\ \text{sen}(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$ respectively and the Euclidean distance is conserved.

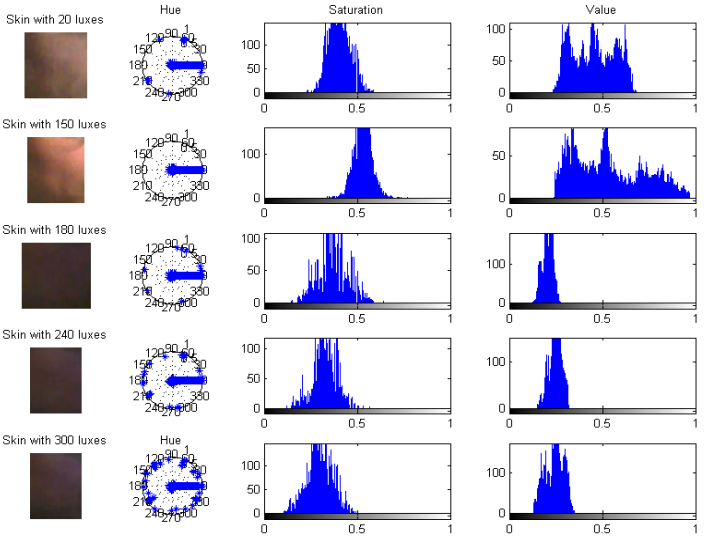


Figure 3 Skin segment described with HSV model and Hue component histograms.

Another important topic in this approach is the object's color, because using the color segmentation permits a wide range of illumination independence and powerful discrimination [Zhuang 2013], this step obtains segments that correspond to searching color in the image. Actually, much of commercial cameras bring a RGB color image with resolution $n_0 \times m_0$, and $q = 256$ quantization levels and three-color components indexed by l ($m = 1$ for red, $n = 2$ for green, $l = 3$ for blue color) represented in equation 1.

$$I(m, n, l) \in \{0 < \mathbb{Z} < 255\}$$

$$\forall \ \{m|0 \leq m \leq m_0\},\{n|0 \leq n \leq n_0\},\{l|1 \leq l \leq 3\} \quad (1)$$

In order to use HSV color scheme the transformation is composed by $H(x,y)$, $S(x,y)$ and $V(x,y)$ defined equations 2-4:

- $if\ MAX(x,y) = MIN(x,y)$

$$H(x,y) = Undefined \quad (2a)$$

- $if\ MAX(x,y) = I(x,y,1)\ \ and\ \ I(x,y,2) \geq I(x,y,3)$

$$H(x,y) = \frac{60°\left(I(x,y,2) - I(x,y,3)\right)}{MAX - MIN} + 0° \quad (2b)$$

- $if\ MAX(x,y) = I(x,y,1)\ \ and\ \ I(x,y,2) < I(x,y,3)$

$$H(x,y) = \frac{60°\left(I(x,y,2) - I(x,y,3)\right)}{MAX - MIN} + 360° \quad (2c)$$

- $if\ MAX(x,y) = I(x,y,2)$

$$H(x,y) = \frac{60°\left(I(x,y,3) - I(x,y,1)\right)}{MAX - MIN} + 120° \quad (2d)$$

$$S(x,y) = \begin{cases} 0, & if\ MAX(x,y) = 0 \\ 1 - \dfrac{MIN(x,y)}{MAX(x,y)}, & otherwise \end{cases} \quad (3)$$

$$V(x,y) = MAX(x,y) \quad (4)$$

Where $MIN(x,y) = \text{argmin}_{1 \leq l \leq 3} I(x,y,l)$ and $MAX(x,y) = \text{argmax}_{1 \leq l \leq 3} I(x,y,l)$.

As can be observed, if constants k_1 and $k_2$define different illuminations of one pixel and $k_1 > k_2$, it is possible to define the pixel $p_1 = (k_1 r, k_1 g, k_1 b)$ as more illuminated than $p_2 = (k_2 r, k_2 g, k_2 b)$ within the RGB color model. After applying the transformation (2)-(4) from RGB to HSV to both pixels, the H component result is $H(p_1) = H(p_2)$. Therefore, the H component does not depend on white illumination. Thus, we can get color segmentation using equation 5.

$$b(x,y) = \begin{cases} 1, \forall\ (\mu_A < H(x,y) < \mu_B) \\ 0, \quad otherwise \end{cases} \quad (5)$$

Where μ_A and μ_B are determined by the color statistics in terms of their distribution function modeled by $N(a,\sigma)$ using $a \pm 3\sigma$ to enclose 88% of its occurrences as specified by Tchebycheff [Box, 2005]. Thus $\mu_A = a - 3\sigma$ and $\mu_B = a + 3\sigma$

After the image is binarized a connected-component labeling process is needed to identify each blob [Sossa, 2016] using equation 6.

$$O_i(x,y) = \text{labeling}\big(b(x,y)\big) \tag{6}$$

Where $O_i(x,y) \cap O_j(x,y) \in \emptyset \quad \forall \quad j \neq i \quad \forall \quad i,j = 1,2,\dots,LN$ and LN is the total number of blobs.

Shape description by Hu moments

Shape is calculated using Hu invariants to define each 2D shape in numerical terms; these are invariant to translations, rotations and scale changes [Hu, 1962]. The use of Hu descriptors generalizes the use of images at different resolutions (typically when an ANN is used, its inputs are connected directly to each pixel to classify objects in images; here this is unnecessary). This helps in the classification stage when describing identical objects of similar quantities in Hu invariant space. Figure 4 compares the behavior of Hu invariants for different type of facial silhouettes.

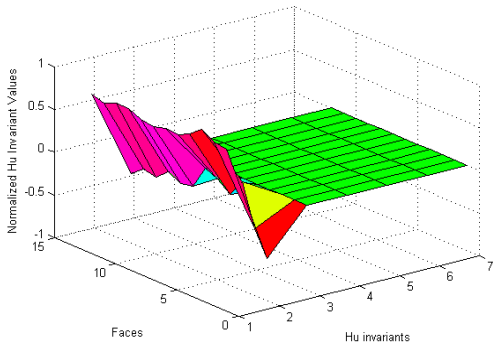


Figure 4 Normalized Hu invariants for 11 different facial silhouettes.

The values of this graphic are normalized with respect to the maximum descriptor value. As can be seen, in all classes the shapes of the graphs are similar; thus, it is convenient to use the DF in order to discriminate between facial sizes.

Dimensional features (DFs) extraction

DF range determination is important when using dimensionality in 3D object recognition. Initially, a DF or DFs must be chosen: hypothetically, a DF for human body classification could be height; in vehicles such as compact cars, vans or trailers, the tire diameter could be used; color, shape, or the distance between the extremes of the thread could be chosen as DFs for screws. The DFs provide the specific dimensional information that is necessary for object recognition. The next step is DF measurement, which can be done directly with any tool or by using the stereo vision system. In order to use the stereo vision system to measure the k-th DF in offline and online stages, the mathematical model parameters of stereo calibrated cameras relate points in 3D space to their respective points in the stereo images. This allows measurement of any object dimensión, equations 7 and 8.

$$\begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix}_k = \text{triangulation}(P_0^L, P_0^R)_k \tag{7}$$

$$\begin{bmatrix} X_1 \\ Y_1 \\ Z_1 \end{bmatrix}_k = \text{triangulation}(P_1^L, P_1^R)_k \tag{8}$$

Finally, to obtain the DF, one must only calculate the distance between two 3D points, equation 9.

$$DF_k = \sqrt{(X_u - X_l)^2 + (Y_u - Y_l)^2 + (Z_u - Z_l)^2} \tag{9}$$

Where $(X_u, Y_u, Z_u)^T{}_k$ and $(X_l, Y_l, Z_l)^T{}_k$ are coordinates of the k-th DF extremes, and P_x^R and P_x^L are the right and left points on the image of the k-th DF extremes, respectively.

ANN training

The classification stage is implemented with 10 neurons in one hidden layer Neural Network (NN), trained to identify the shape of objetcs using Hu invariants (equation 10). Its architecture consists in fourteen inputs connected to a “tansig” neuron and a “logsig” function at the output.

$$O_i^s = NN\left(\phi_{1,i}^s, \phi_{2,i}^s, \phi_{3,i}^s, \phi_{4,i}^s, \phi_{5,i}^s, \phi_{6,i}^s, \phi_{7,i}^s\right) \tag{10}$$

Where O_i^S has binary values, 0 to indicate that is not a indicated object and 1 to express that it is.

3. Results

In this section we present four experiments that show the potential of 3DNOD-CSD. The first experiment demonstrates the discrimination of dolls' faces from human faces in stereo photos. The second example shows the differentiation of a human face and its photo in an image. The third experiment demonstrates discrimination between similar Mexican coins. The fourth experiment presents the detection of utility poles in real conditions.

Doll's face detection in stereo images

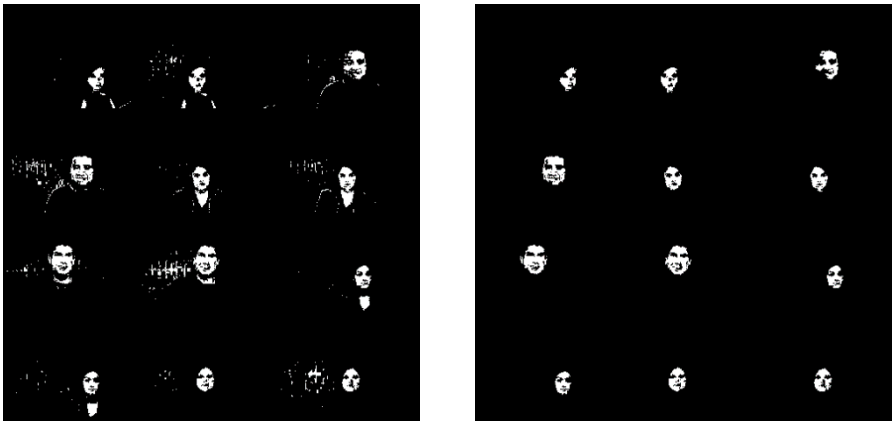
For this experiment we used 15 photos of six different faces captured in stereo mode at different distances from the cameras. A conventional facial recognition system detects the faces of the dolls as human faces. This experiment reveals how our method identifies the dolls' faces as such by using their color, shape and a specially chosen dimension.

Figure 5 illustrates the six different faces used in the experiment, including the RGB color version and the color-segmented images used in order to get the human skin segments or blobs. The Hu invariants of all segments are computed to describe their shapes. Finally, the ANN classifies face shapes from all blobs.

Next, a calibrated stereoscopic camera system is used to measure the DF. The chosen feature is facial height as it will differentiate the doll's and the person's skull sizes; furthermore, it is very easy to obtain automatically. The heuristics to obtain the facial height is based on the highest and lowest pixel on the "Y" axis from the binary silhouettes selected by Hu invariants.

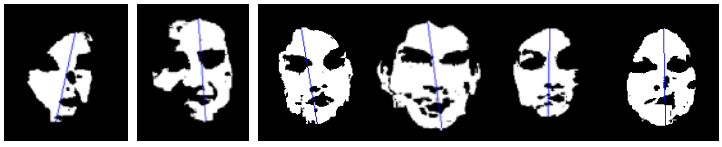
To make a detailed analysis of each shape the objects are isolated in an image that holds the original position information. Statistical features of the faces are ready to be compared with the database of features to differentiate between human and doll faces. The DF approach is shown in figure 6 and compared in Figure 7, where it is very clear that the doll face is approximately 47 millimeters and the human faces are between 140 and 180 millimeters.

a) RGB images. *b) Color segmented images.*

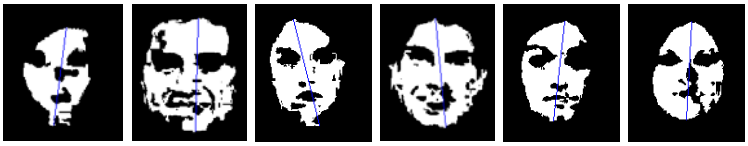


c) Binarized version of color segmented image. *d) Shape detected on images.*

Figure 5 Objects used in experiments.



a) Left Camera Images.



b) Right Camera Images.

Figure 6 Dimensional feature of facial height for doll and human faces.

One can see that the faces represented in terms of neural network and Hu invariants are similar, but when facial height is calculated the doll is easily discriminated, even if it has an identical skin color. Figure 7 presents the different heights of the faces and the easy discrimination between the doll and human faces.

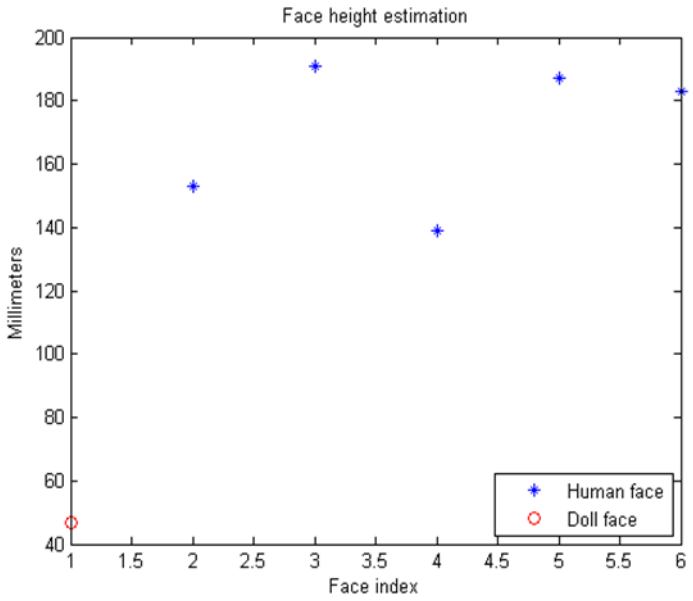


Figure 7 Dimensional feature of facial height for doll and human faces.

Distinction between in situ face and photograph of face

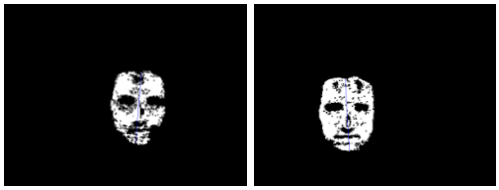
With a typical facial recognition system, a photograph of a face is detected as a real face. This experiment demonstrates how 3DOD-CSD automatically distinguishes between a photo taken of a person in situ and a photograph of the same person's face. 3DOD-CSD begins with Gaussian color filtering to locate the color segments of the object to be found, as shown in figure 8. Next, the trained neural network detects facial silhouettes using Hu invariants as shape descriptors. Finally, the maximum distance of the facial blob is calculated with stereo triangulation.

Automatic visual recognition of Mexican coins

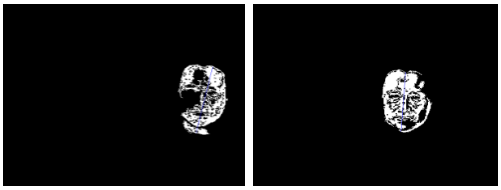
In this experiment 3DOD-CSD detects and discriminates between three Mexican coins, illustrating the robustness and advantages of using 3DOD-CSD in an uncontrolled environment. The coins are of three different sizes but one face of each of the three coins has an identical color and image. Therefore, it is crucial for the object detection system to be able to discriminate between their sizes when the identical faces are presented. Figure 9 shows the 1, 2 and 3 peso coins; note that one side displays their value in pesos while the other side has the identical image of an eagle with a snake in its beak.

a) Source image.

b) Color segmented image.



c) Shape detected by Hu invariants.



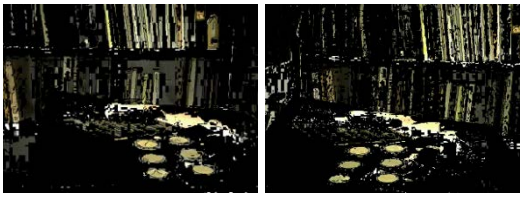
d) Shape detected by Hu invariants.

Figure 8 Discrimination between in situ face and photograph of face.

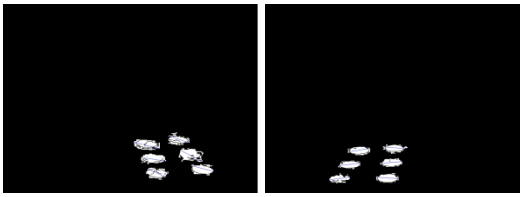
Figure 9 Mexican coins: one face has the value in pesos.

In figure 10 one can see that the coins are presented in an orientation in which the coins appear to be of identical size. The background is multihued. In figure 10 the six coins appear to a human viewer to be identical. Row (a) presents the color images, (b) the color-segmented image, using the HSV model, (c) the coins' blobs, with the maximum axis of each blob to the left to calculate the radius, correlating with the corresponding points in the right image. Figure 11 displays the measured radius.

a) Color images.



b) HSV color-segmented image.



c) blobs.

Figure 10 Mexican coins in busy scene.

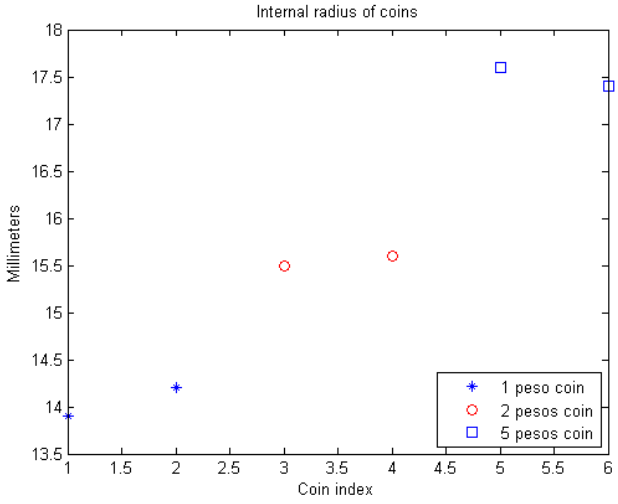


Figure 11 Internal radius of six coins in Figure 10 measured with a stereo vision system.

One can see that 3DOD-CSD discriminates between the coins even when they are presented in an orientation in which a human viewer might be mistaken.

Utility pole detection

This experiment shows the ability of 3DOS-CSD to detect utility poles in images taken in an uncontrolled outdoor environment. The following experiment uses a

database of 200 different photos of utility poles, taken outdoors at differing distances between the camera and the target, 100 photos each of left and right sides. Figure 12 shows the images taken with conventional RGB cameras at 432x576 pixel resolution.

Figure 12 Left (up row) and right photos of utility poles used in different conditions.

Figure 13 presents the segmented images of the utility poles. The color filter in this case for the H component of the HSV model is $\bar{x} = 0.585$ and $\sigma = 0.045$. The neural network training stage used 40 photos with all segments that include at least one UP and many non- UP segments.

a) Original potos.

b) Color-segmented images.

c) Segments identified by Hu moments and Neural Network.

Figure 13 Photos source: The authors.

4. Discussion

One can see that 3DOD-CSD differentiates the utility poles within an uncontrolled environment. The proposed 3DOD-CSD visual object detection system uses color,

shape and dimensional features to detect objects of interest. A stereo vision system is required for hardware and Gaussian color segments, Hu invariants descriptors, and an ANN for software. 3DOD-CSD does not address the problem of object recognition directly but recognizes global color, shape, and dimensional differences. Camera sensitivity and compensation systems determine the robustness of illumination sensing, although even a genius eye312 conventional camera obtains a wide work range of between 20 to 400 luxes. As the H component of the HSV model is independent of the V component describing the quantity of white illumination, the H component provides stability in terms of illumination. Table 1 displays the advantages of the 3DOD-CSD system.

5. Conclusion

This paper presents an innovative method of automatic object detection based on three common features to classify in human vision: color, shape, and dimensions. Experimental results indicate that the method is able to detect objects using HSV color segmentation, Hu invariants as descriptors and an ANN in the classification stage. With the help of a calibrated stereo vision system, it is possible to differentiate similar objects of different sizes. This visual recognition system is flexible in terms of uses and can be implemented at low cost. By other hand, the lighting range depends of the camera sensitivity. In this work using a 312 eye camera the range of work is between 20 and 4000 luxes with a high rate of correct detection compared with the method reported by [Zalevsky, 2007].

Table 1 Comparison of three methods of color object detection

	Barranco et. All.	Kviatkovsky et. All (2013)	Zhuang et. All (2012)
Shape description	Yes	No	No
Scale invariance	Yes	Yes	Yes
Translation invariance	Yes	Yes	Yes
Position respect to the camera invariance	Yes	Yes	Yes
Neurons Quantity	8 at input 10 at Hidden Layer 1 output	No used	One per pixel
Color description	Yes	Yes	Yes
Dimensions description	Yes	No	No
Illumination range	20 -400 Luxes	No specified	No specified

Acknowledgments

The authors greatly appreciate the support of CONACyT, TecNM and PRODEP. We would also like to thank Laura Heit for her valuable editorial help.

6. References

[1] Amos A., Suppa M., Gerth W. (2001). Detection of stair dimensions for the path planning of a bipedal robot, International Conference on Advanced Intelligent Mechatronics Proceedings. 1291-1296. DOI: 10.1109/AIM.2001.936909.

[2] Baek H. S., Choi J. M., Lee B. S. (2010). Improvement of distance measurement algorithm on stereo vision system (SVS), Proceedings of the 5th International Conference on Ubiquitous Information Technologies and Applications (CUTE), 1-3. DOI: 10.1109/ICUT.2010.5678176.

[3] Barranco G. A. I., Medel J. J. J. (2009). Digital Camera Calibration Analysis Using Perspective Projection Matrix, in Proceedings of the 8th WSEAS International Conference on Signal Processing, Robotics and Automation, 321-325. Available at: http://dl.acm.org/citation.cfm?id=1558971.

[4] Barranco G. A. I., Medel J. J. J. (2011). Automatic object recognition based on dimensional relationships, Computación y Sistemas, 15(2), 267-272. Available at: http://www.scielo.org.mx/pdf/cys/v15n2/v15n2a11.pdf.

[5] Barranco G. A. I., Martínez D. S., Gómez T. J. L. (2013). An Approach for Utility Pole Recognition in Real Conditions, PSIVT Workshop on Quality Assessment and Control by Image and Video Analysis, 113-121. DOI: 10.1007/978-3-642-53926-8_11.

[6] Box E., Junter S., Hunter W. (2005) Statistics for experimenters, design, innovation and discovery, USA, John Wiley & Sons Inc.

[7] Costa M. S., Shapiro L. G. (2000). 3D Object Recognition and Pose with Relational Indexing, Computer Vision and Image Understanding, 79(3), 364–407. DOI: 10.1006/cviu.2000.0865.

[8] Evangelidis G. D., Hansard M., and Horaud R. (2015). Fusion of Range and Stereo Data for High-Resolution Scene-Modeling, IEEE Transactions on

Pattern Analysis and Machine Intelligence, 37(11), 2178 - 2192. DOI 10.1109/TPAMI.2015.2400465.

[9] Forsyth D., Mundy J.L., Zisserman A., Coelho C., Heller A., Rothwell C. (1991). Invariant Descriptors for 3D Object Recognition and Pose, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13(10), 971 – 991. DOI: 10.1109/34.99233.

[10] Gonzalez R. C., Woods R. E. (2006) Digital Image Processing, USA, Prentice Hall.

[11] Gupta M., Agrawal A., Veeraraghavan A., Narasimhan S.G. (2011). Structured light 3D scanning in the presence of global illumination, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 713-720. DOI: 10.1109/CVPR.2011.5995321.

[12] Hartley R., Zisserman A. (2003). Multiple view geometry in computer vision, UK, Cambridge.

[13] Hu M. K., (1962). Visual Pattern Recognition by Moment Invariants, IRE Transactions on Information Theory, vol. IT-8, 179–187, DOI: 10.1109/TIT.1962.1057692.

[14] Hu Y., Jiang D.; Yan S., Zhang L., Zhang H. (2004). Automatic 3D reconstruction for face recognition, Sixth IEEE International Conference on Automatic Face and Gesture Recognition Proceedings, 843–848. DOI: 10.1109/AFGR.2004.1301639.

[15] Linderberg T., Garding T. (1997). Shape-adapted smoothing in estimation of 3-D shape cues from affine deformations of local 2-D brightness structure, Image and Vision Computing, 15(6), 415–434, Available at: http://www.sciencedirect.com/science/article/pii/S026288569701144X

[16] Marquez M. A., Vargas C. A., and Arguello H. (2016). Automatic detection of bumblebees using video analysis, Ingeniería e Investigación, 36(3), 81-84. DOI: http://www.revistas.unal.edu.co/index.php/ingeinv/article/view/54267/58067.

[17] Mustafah Y. M., Noor R., Hasbi H., Azma A. W. (2012). Stereo Vision Images Processing for Real-time Object Distance and Size Measurements,

International Conference on Computer and Communication Engineering (ICCCE 2012), 659 – 663. DOI: 10.1109/ICCCE.2012.6271270.

[18] Kviatkovsky I., Adam A., Rivlin E., (2013). Color Invariants for Person Reidentification. IEEE Transactions on Pattern Analysis and Machine Intelligence. 35(7), 1622-1634. DOI: 10.1109/TPAMI.2012.246.

[19] Ocegueda O., Fang T., Shah S., Kakadiaris I. (2013). 3D Face Discriminant Analysis Using Gauss-Markov Posterior Marginals, IEEE Transactions on Pattern Recognition and Machine Intelligence, 35(3), 728-738. DOI: 10.1109/TPAMI.2012.126.

[20] Osman M.K. (2009). 3D object recognition using MANFIS network with orthogonal and non-orthogonal moments, 5th International Colloquium on Signal Processing & Its Applications, 302 – 306. DOI: 10.1109/CSPA.2009.5069239.

[21] Pajares G., De la Cruz J. M. (2004). Visión por computador imágenes digitales y aplicaciones, España, AlfaOmega.

[22] Rahman K. A., Hossain M. S., Bhuiyan M. A., Tao Z., Hasanuzzaman M., Ueno H. (2009). Person to Camera Distance Measurement Based on Eye-Distance, 3rd International Conference on Multimedia and Ubiquitous Engineering (MUE'09), 137-141. DOI: 10.1109/MUE.2009.34.

[23] Zhuang H., Low K. S., Yau W. Y. (2013). Multichannel Pulse-Coupled-Neural-Network-Based Color Image Segmentation for Object Detection. IEEE Transactions on Industrial Electronics, 59(8), 3299 – 3308. DOI: 10.1109/TIE.2011.2165451.

[24] Sigal L., Sclaroff S., Athitsos V. (2004). Skin Color-Based Video Segmentation under Time-Varying Illumination, IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(7), 862 – 877, DOI: 10.1109/TPAMI.2004.35.

[25] Sossa H. (2006) Rasgos descriptores para el reconocimiento de Objetos, Instituto Politécnico Nacional, 113-176.

[26] Trucco E., Verri A. (1998) Introductory techniques to 3D computer vision, USA, Prentice Hall.

[27] Ren H., Zhong Q., Kang J. (2009). Object recognition algorithm research based on variable illumination, IEEE International Conference on Automation and Logistics (ICAL'09), 1609–1613. DOI: 10.1109/ICAL.2009.5262717.

[28] Wang J., Athtsos V., Sclaroff S., Betke M. (2008). Detecting Objects of Variable Shape Structure with Hidden State Shape Models, IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(3): http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=4359323.

[29] Xiaoming L., Tian Q., Wanchun C., Xingliang Y. (2009). Real-Time Distance Measurement Using a Modified Camera, IEEE Sensors Applications Symposium, 54-58. 10.1109/SAS.2010.5439423.

[30] Yi-Xing L., Ying L., Yu G., Li-Tao K., Xiao-Qi C., Xiao-you S. (2007). Features of human skin in HSV color space and new recognition parameter, Optoelectronics Letters, 3(4), 312-314, DOI: 10.1007/s11801-007-6175-3.

[31] Zhang Z. (2000). A Flexible New Technique for Camera Calibration, IEEE transactions on pattern analysis and machine intelligence, 22(11), 1330-1334: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=888718.

[32] Zalevsky Z., Shpunt A., Maizels A., Garcia J. (2007). Method and system for object reconstruction, World Intellectual Property Organization: http://www.google.com/patents/US20100177164.