Computación y Sistemas Vol. 6 No.3 pp. 227 - 232
© 2003, CIC - IPN. ISSN 1405-5546 Impreso en México

PH. D. THESIS ABSTRACT

Methodologies for Reducing the Amount of Required Images Used for Articled-Object Recognition

Metodologías para la Reducción del Número de Imágenes Requeridas para el Reconocimiento de Objetos Articulados

Graduated: Luis Carlos Altamirano Robles
Graduated on March 19, 2002
Instituto Mexicano del Petróleo, PIMAyC
Eje Central Lázaro Cárdenas 152, C.P. 07730
Del. Gustavo A. Madero, México, D. F.
e-mail: laltamir@imp.mx
Advisor 1: Leopoldo Altamirano
Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México
e-mail: robles@inaoep.mx
Advisor 2: Matías Alvarado Mentado
Instituto Mexicano del Petróleo
e-mail: matiasa@cic.ipn.mx

Abstract

The appearance-based approaches are such that any object's model is made through a set of training images describing the object's appearance. In this PhD. thesis, the usage of non-uniform sampling is introduced for building this image set. Non-uniform sampling is held by a linear interpolation technique, which is used to determine the strictly necessary images. Main results are: a significant reduction in the quantity of necessary images for the object's model, as well as more precise models than those obtained by uniform sampling. Non-uniform sampling is used in conjunction with the eigenspaces technique for object recognition, producing a more efficient hybrid technique.

Keywords: non-uniform sampling, object recognition, appearance-based model, interpolation, eigenspaces.

Resumen

Los enfoques basados en apariencia construyen el modelo de un objeto, por medio de un conjunto de imágenes de entrenamiento que describe la apariencia del objeto. En esta tesis doctoral se propone el empleo del muestreo no-uniforme para generar tal conjunto de imágenes. El muestreo no-uniforme es soportado mediante una técnica de interpolación, que determina cuáles son las imágenes estrictamente requeridas para construir el modelo. Los resultados principales obtenidos con esta propuesta son: una reducción significativa de la cantidad de imágenes requeridas para construir el modelo del objeto, además de una mejora en la precisión de los modelos así generados, respecto a los obtenidos con muestreo uniforme. El muestreo no-uniforme es empleado junto con la técnica de espacios fundamentales (*eigenspaces*) para realizar el reconocimiento del objeto, obteniéndose una técnica híbrida más eficiente.

Palabras Clave: non-uniform sampling, object recognition, appearance-based model, interpolation, eigenspaces.

1 Introduction

Geometrical approaches have been widely used in computer vision for object recognition. However, they have two main disadvantages: segmentation errors and loss of information contained in the image caused by the segmentation. Appearance-based approaches were proposed for object recognition, as an alternative for the geometrical approaches, to avoid such disadvantages (Shapiro and Costa, 1994).

Several appearance-based methods have been proposed in the literature. The simplest method is the correlation of an image with an object template. The so-called eigenspace approach was introduced in (Turk, 1991; Murase and Nayar, 1995). Patch detection was employed in (Nelson and Selinger, 1997). Neural nets have also been explored (Pauli *et al.*, 1995; Poggio and Beymer, 1996). Probabilistic techniques were also introduced (Moghaddam and Pentland, 1996; Pope and Lowe, 1996). Several methods have been proposed in the literature for object recognition without previous segmentation (Ohba and Ikeuchi, 1996; Schiele and Crowley, 2000).

Appearance-based techniques use an image set taken from the objects to be processed, and they build the corresponding models using the image set. The acquisition of the image set is usually carried out by means of a discrete and uniform sampling of images (equally spaced), taken around the object, and independent of object's characteristics (Stevens and Beveridge, 2000), so, for any object, the quantity of images is supposed to be an independent constant of the object type.

However, image acquisition could be object-dependent. This can be observed while considering extreme examples of objects that require only one image to build their model, for example, a homogeneously colored sphere. In contrast, there are objects of great complexity in their appearance, which need a large number of images. Current techniques do not take into account this information and, therefore, such techniques require a constant time to build the model of any object, without depending on the object complexity.

So far in the current literature, efforts towards reduction of the quantity of characteristic views have been made (Epstein *et al.*, 1995; Belhumeur and Kriegman, 1998). However, if viewing position (instead illumination) is changing, no characterization is done. For working on view position changes of the object, diverse techniques on image synthesis have been introduced. However, there are several differences between the appearance-based approach and image synthesis.

Several techniques for modeling objects with less views have been introduced for object recognition (Koenderink and Van Doorn, 1979; Murase and Nayar, 1995), however, they are computationally expensive and hard to use; recently, have been presented some works to determine how many and which images are necessary to model an object. However, their application is restricted to specific objects as faces (Cootes *et al.*, 2000), or they only work with the object's shape (Mokhtarian and Abbasi, 2000).

We propose the use of non-uniform sampling to reduce the quantity of necessary images for object modeling and recognition. Non-uniform sampling is supported by a simple proposed technique, which determines the required images to approximate the object's appearance within a given error (ε). A linear interpolation and SSD-distance criterion guides the images election.

Non-uniform sampling has been widely used in diverse areas, as approximation theory or image processing and image synthesis, but not in the context of appearance-based object recognition. Although interpolation has been widely used in digital image interpolation techniques, we propose the usage of interpolation as criterion of selection, instead of using it only as a technique to obtain interpolated images. Object recognition is made by using the eigenspaces technique to show the usage of the proposed technique. Experimental results show that reduction in the quantity of required views can be possible, if the behavior of the object's appearance is taken into count, and no pose estimation is necessary.

2 Preliminaries

2.1 Interpolation

Definition 1: A trajectory in $\mathbf{R}^n$ is a function $\sigma:[a,b]\rightarrow\mathbf{R}^n$. The points $\sigma(a)$ and $\sigma(b)$ are trajectory's extremes.

Definition 2: Let σ_1: $[a, b]\rightarrow\mathbf{R}^n$ be a trajectory. Let $C=\{(x_1,y_1), (x_2,y_2),\ldots,(x_m,y_m)\}$ be such that $\sigma_1(x_i)=y_i$, for $i=1,\ldots,m$, with $x_1\leq x_2\leq\ldots\leq x_m$. We say that a trajectory $\sigma_2:[a,b]\rightarrow\mathbf{R}^n$, interpolates $\sigma_1(x)$ inside the interval $[x_1,x_m]$, if $\sigma_2(x_i)=y_i$, for $i=1,\ldots,m$.

Notice that $\sigma_2(x)$ is not uniquely determined. For this reason, for an ε given, $\sigma_2(x)$ is selected such that $|\sigma_1(x) - \sigma_2(x)| < \varepsilon$, for all x in $[x_1,x_m]$. Such ε is named the associated interpolation error. If the trajectory $\sigma_2(x)$ that interpolates $\sigma_1(x)$ is a straight line, interpolation is linear; otherwise, it is non-linear.

Sometimes, it is impossible to find a trajectory $\sigma_2(x)$ that meets the error criterion given inside the interval. In this case, the interpolation problem of $\sigma_1(x)$ is changed for the $\sigma_1(x)$ *piecewise interpolation problem*: it is necessary to find a set of m-1 trajectories $\sigma_2(x)$, $\sigma_3(x)$,..., $\sigma_m(x)$, that interpolate $\sigma_1(x)$, respectively, in the intervals $[x_1,x_2],[x_2,x_3],\ldots[x_{n-1},x_m]$, within the tolerance ε.

Definition 3: Let $\mathbf{X}=(x_1, x_2, ..., x_n)$ and $\mathbf{Y}=(y_1, y_2, ..., y_n)$ be two vectors in $\mathbf{R}^n$. The SSD distance (sum-of-squared-difference) between $\mathbf{X}$ and $\mathbf{Y}$ is defined by:

$$\|\mathbf{X}-\mathbf{Y}\|^2 = \sum_{1}^{n}(x_i - y_i)^2$$

2.2 Eigenspaces Overview

The eigenspaces technique (Murase and Nayar, 1995) comprises a set $\mathbf{I}$ of training images into a compact representation that captures the key appearance characteristics of the object. Although uniform sampling was used in (Murase and Nayar, 1995), the set $\mathbf{I}$ can be a uniform or non-uniform sampling of the visual workspace. This compression process is computationally intensive, and its complexity depends on the size of $\mathbf{I}$ (if the size of $\mathbf{I}$ is smaller than the number of pixels that constitutes the images, otherwise, depends upon the number of pixels). For this reason, reducing the size of $\mathbf{I}$ is desired, as it was proposed in (Altamirano *et al.*, 2003) by means of non-uniform sampling.

3 Image Acquisition

In order to generate appearance-based models of an object, we begin with the object image acquisition. By taking images around the object, the acquisition is carried out. The object rotates in front of the camera (or the camera rotates around the object) for an angle θ_i, respect to a fixed initial position to take the i^{th}-image $\mathbf{I}_i$, with $i=1,.., r$ and $\theta_{i-1} < \theta_i$. The camera and the illumination source can be moved around the object, before object modeling. However, when the image acquisition process begins, both positions must stay fixed in the rest of image acquisition process and in the recognition phase. In addition, the background of images acquired to model the object is usually controlled.

The images acquired in the upper half sphere $\mathbf{I}_1,\ldots,\mathbf{I}_r$ are represented as nxm matrixes, and matrix's columns are stacked to form vectors $\mathbf{v}_1,\ldots,\mathbf{v}_r \in \mathbf{R}^{nm}$. Looking at the

Definition 1, we can see that such vectors belong to a trajectory $\sigma:[0,2\pi]\rightarrow \mathbf{R}^{nm}$ parameterized for the angle θ of the object rotation in front of the sensor, regarding a fixed initial position. This trajectory is called in the literature **the trajectory determined by object's appearance**, or simply, **object's appearance.**

Usually, proposed techniques for object recognition assume for simplicity, that $\theta_i - \theta_{i-1} = k$, with k a constant, which means that the object always rotates to the same angle (with respect to the current position), in order to capture the next image. Typical values found in the literature for k oscillate between 5° and 12°. These values are based on empiric observations, and therefore do not present any formality to set them. Moreover, this value is fixed and independent of the object class (texture, shape, color, etc.).

Acquiring images in a uniform way supposes a large storage capacity and required time for model calculation.

4 Non-Uniform Image Acquisition

To overcome the previously commented limitations, we propose to use non-uniform sampling for modeling the object's appearance (Altamirano *et al.*, 2003). Non-uniform sampling is supported through of a simple proposed technique that adapts the value of the angles θ_i according to the object's appearance and precision requirements.

This technique is based on the observation that object's appearance can be approximated by means of piecewise linear interpolation (see Definition 2), within an error ε. Of course, it is possible to create another techniques to support non-uniform sampling, probably, using non-linear interpolation, but we propose this technique as an easy way of exploring non-uniform sampling.

The technique starts with the acquisition of a couple of images denoted by **X**, **Y**, taken at 0° and 180°, respectively. Next, the technique interpolates linearly between **X** and **Y** using the function:

$$g(\lambda) = \lambda\mathbf{X} + (1-\lambda)\mathbf{Y};\quad 0 \leq \lambda \leq 1.$$

Later, we take another image **Z** in the middle point (90°), and compare **Z** to the image **Z'** obtained by linear interpolation at the same point, which corresponds to g(0.5). If SSD (see Definition 3), between **Z** and **Z'** is smaller than ε, the algorithm determines that only a couple of images are required to approximate object's appearance between 0° and 180°. If this condition is not fulfilled, the interval is divided into 2 parts: the first one contains the images between 0° and 90° whereas another part includes the images between 90° and 180°. In a similar way an analysis is made for the images included between 180° and 360°. The process is iterative analyzing each new interval generated. The algorithm finishes when it is possible to approximate the object's appearance within an error ε. A complete description of the algorithm can be found in (Altamirano *et al.*, 2003).

4.1 Automatic Determination of Precision

The proposed technique requires a precision value ε, to start, but the precision that should be used for each particular object cannot be known *a priori*. So, in order to the system automatically suggests a precision value, some strategy is claimed. In (Altamirano *et al.*, 2003), we introduce two statistical strategies for the system that automatically determines one value for this precision. These strategies will work alongside as the proposed technique does. Both strategies are based on the observation that the k first images acquired by the proposed technique can be used to obtain information, about local variations in object's appearance. Such information can be used to estimate a precision value. The first one approximates the average of k samples, and the second one approximates the maximum precision using the physical limit of the rotation system. In both cases, estimation will be best if k is bigger.

4.2 Generalized Non-Uniform Sampling

In previous Sections, we supposed that the object is in the turntable center, and the object rotates in front of the camera. Moreover, we supposed that camera was in a fixed position with respect to turntable, and stays static. Finally, we supposed that the illumination source was fixed with respect to the camera and the turntable. However, we can notice that if are removed any or all of these suppositions, the object's appearance would be defined for several parameters, and the exposed technique in Section 3 cannot applicable anymore. Some parameters could be considered (but not restricted to) include: Turntable angle, Camera elevation angle (relative to turntable plane), Position of illumination source, Position of any part of an articulated object, etc. We proposed the generalization for working on N parameters of the basic technique exposed in Section 3. Of course, because we are considering N parameters, several changes are claimed. A complete description of this generalized case can be found in (Altamirano *et al.*, 2002).

4.3 Non-Uniform Sampling and Eigenspaces

We mentioned that the technique of eigenspaces is computationally intensive, and thereby, the proposed technique can be used to reduce the required time for the model calculation, improving also its precision. The technique of eigenspaces can use the determined images by the proposed algorithm as its training set of images, instead of the larger number used currently. Notice that it is only true, if pose estimation doesn't matter, but the proposed technique can be easily modified, to verify good approximation in both parameter and appearance spaces. Of course, in this case, image quantity reduction may be not significant. However, in several applications where pose estimation no matters, the computation of eigenspaces will be faster using a non-uniform sampling.

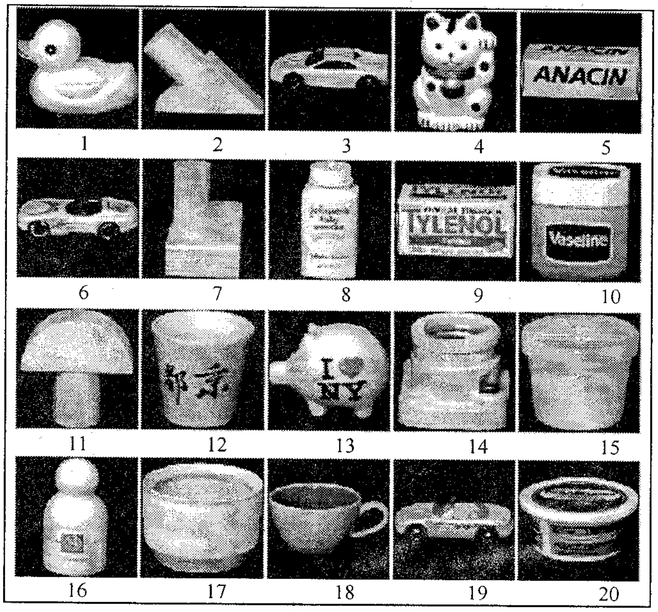

Figure 1: Some objects taken from Columbia Object Image Library (Nayar *et al.*, 1996) used to study non-uniform sampling

5 Experimental Results

A software system that acquires the necessary images for modeling the appearance of objects was developed. The software was coupled to a motorized turntable that rotates the object. The rotation system received the commands of the software in order to acquire images in the positions determined by the algorithm. The software system determines how many and which images are required to satisfy the precision criterion ε, for each object studied. The number of images founded by the proposed technique represents a significant reduction with respect to traditional approaches. Of course, this number will correspond to required precision for the object model. In addition to the number of images, the algorithm determines *which* images are required for each particular object, and for each value of the error ε. To experiment the proposed technique, only two parameters were used: turntable angle and camera elevation angle.

The proposed technique was applied to a large image base (about 50 objects); 20 objects of it are shown in Figure 1. In Table 1 are shown experimental results for objects in Figure 1 (only one parameter: turntable rotation). Notice the significant reduction achieved with non-uniform sampling respect to uniform sampling.

Object number	Maximum precision (36 images)	Number of necessary images	
		Uniform sampling	**Non-Uniform sampling**
1	2073	36	**32**
2	3514	36	**27**
3	5103	36	**18**
4	2945	36	**26**
5	5252	36	**23**
6	7282	36	**18**
7	2360	36	**31**
8	2152	36	**25**
9	5817	36	**26**
10	3287	36	**22**
11	2509	36	**29**
12	2319	36	**11**
13	4164	36	**22**
14	3448	36	**26**
15	1445	36	**6**
16	1267	36	**13**
17	2733	36	**6**
18	2545	36	**26**
19	4655	36	**18**
20	3647	36	**9**

Table 1: Comparison between uniform and non-uniform sampling.

6 Conclusions

Non-uniform sampling was introduced as a fast way for building appearance-based models when pose estimation doesn't matter. A simple technique has been presented for non-uniform image acquisition of an object; the technique is based upon linear interpolation and is used to determine the strictly necessary images to capture object's appearance, within a defined error ε, and to diminish the buffered images. Also it was shown how the technique can be used together with the eigenspaces technique, for reduction of computational time used in model construction.

Acknowledgments

This research was sponsored in part by CONACyT under Ref. 131991A. L.C. Altamirano was supported by CONACyT graduate scholarship No. 89635. The author would like to thank Dr. Matías Alvarado from Instituto Mexicano del Petróleo (PIMAyC), for reviewing and correcting a preliminary version of this work.

References

L.C. Altamirano-Robles. "Metodologías para la reducción del número de imágenes requeridas, para el reconocimiento de objetos articulados". Dissertation, PhD. Thesis, Centro de Investigación en Computación, Instituto Politécnico Nacional, México, D.F., March, 2002.

L.C. Altamirano, L. Altamirano, and M. Alvarado. "Generation of N-parametric appearance-based models through non-uniform sampling". Submitted to Pattern Recognition Letters, The Netherlands, 2002.

L.C. Altamirano, L. Altamirano, and M. Alvarado. "Non-Uniform Sampling For Improved Appearance-Based Models". Pattern Recognition Letters, Vol. 24, Issue 1-3, pp. 521-535, The Netherlands, January 2003.

P.N. Belhumeur and D.J. Kriegman. "What is the Set of Images of an Object Under All Possible Ilumination Conditions?". International Journal of Computer Vision, Vol. no. 28, Issue No. 3, 245-260, 1998.

T.F. Cootes, G.V. Wheeler, K.N. Walker and C.J. Taylor. "Coupled-View Active Appearance Models". The Eleventh British Machine Vision Conference, University of Bristol, September 2000.

R. Epstein, P. Hallinan, and A. Yuille. "5±2 Eigenimages suffice: An empirical investigation of low-dimensional lighting models". In proceedings of IEEE Workshop on physics-based modeling in computer vision, 1995.

J.J. Koenderink and A.J. Van Doorn. "The internal representation of solid shape with respect to vision". Biological Cybernetics, Vol. 32, 1979.

B. Moghaddam and A. Pentland. "Probabilistic visual learning for object representation". in Early Visual Learning, Edited by Shree K. Nayar and Tomaso Poggio, New York Oxford, Oxford University Press, 1996.

F. Mokhtarian and S. Abbasi. "Automatic Selection of Optimal Views in Multi-view Object Recognition". The Eleventh British Machine Vision Conference, University of Bristol, September 2000.

H. Murase and S.K. Nayar. "Visual learning and recognition of 3-D objects from appearance". International Journal of Computer Vision, Vol. 14, No. 1, 5-24, January 1995.

S.K. Nayar, H. Murase and S. Nene. "Parametric Appearance Representation". In Early Visual Learning, Edited by Shree K. Nayar and Tomaso Poggio, New York Oxford, Oxford University Press, 1996. Workshop, 1197-1205, 1997.

R.C. Nelson and A. Selinger. "Experiments on (Intelligent) Brute Force Methods for Appearance-Based Object Recognition". DARPA Image Understanding.

K. Ohba and K. Ikeuchi. "Recognition of the Multi Specularity Objects using the Eigen-Window". Proceedings of International Conference on Pattern Recongnition, August 1996.

J. Pauli, M. Benkwitz and G. Sommer. "RBF Networks Appearance-Based Object Detection". Proceedings of ICANN, Paris, Volume 1, pp. 359-364, 1995.

T. Poggio and D. Beymer. "Regularization Networks for Visual Learning". In Early Visual Learning, Edited by Shree K. Nayar and Tomaso Poggio, New York Oxford, Oxford University Press, 1996.

A. Pope and D. Lowe. "Learning Probabilistic Appearance Models for Object Recognition". In Early Visual Learning, Edited by Shree K. Nayar and Tomaso Poggio, New York Oxford, Oxford University Press, 1996.

B. Schiele and J.L. Crowley. "Recognition without Correspondence using Multidimensional Receptive Field Histograms". International Journal of Computer Vision 36(1), 31-50, 2000.

L.G. Shapiro and M.S. Costa. "Appearance-Based 3D Object Recognition". Object Representation in Computer Vision, Vol. I, in Proceedings of International NSF-ARPA Workshop, New York City, NY, USA, Dec. 1994. 51-64.

M.R. Stevens and J.R. Beveridge. "Integrating graphics and vision for object recognition". Kluwer Academic Publishers, 2000.

M. Turk and A. Pentland. "Eigenfaces for Recognition". J. Cognitive Neuroscience, Vol. 3, no. 1, 71-86, 1991.

Luis Carlos Altamirano Robles, *obtuvo el grado de Licenciado en Computación de la Benemérita Universidad Autónoma de Puebla, México, en 1995, y el grado de Maestro en Ciencias de la Computación en 1997, por la misma Institución. En 2002 obtuvo el grado de Doctor en Ciencias de la Computación por el Centro de Investigación en Computación del Instituto Politécnico Nacional. Desde Julio de 2002 es investigador en el Instituto Mexicano del Petróleo, en México, D.F. Sus áreas de interés son: Visión Computacional, Reconocimiento de Patrones, Inteligencia Artificial y Aplicaciones a la Industrial del Petróleo.*

Leopoldo Altamirano Robles, *obtuvo el grado de Licenciado en Computación de la Benemérita Universidad Autónoma de Puebla, México, en 1985. Obtuvo el grado de Maestro en Ciencias en Ingeniería Eléctrica con especialidad en Computación en 1991, en el CINVESTAV-IPN, México. En 1996 obtuvo el grado de Doctor en Ciencias por la Universidad Técnica de Munich, Alemania. Desde 1997 es investigador en el Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México. Sus áreas de interés son: Visión Computacional, Procesamiento de Imágenes, Sistemas Multiagente.*

José Matías Alvarado Mentado, *es Licenciado en Matemáticas por la Universidad Nacional Autónoma de México. En 1998 obtuvo el grado de Doctor en Computación (Excelente Cum Laudae) con especialidad en Inteligencia Artificial por la Universidad Politécnica de Cataluña, España. Fue investigador de tiempo completo desde 1998 hasta 2001 en el CIC-IPN. Desde 2001 es investigador de tiempo completo en el Instituto Mexicano del Petróleo, en México, D.F., Sus áreas de interés son: Sistemas Multiagente, E-negocios, Visión Computacional, Reconocimiento de Patrones y Aplicaciones a la Industria del Petróleo.*