A NEW OBJECT RECOGNITION SYSTEM

Nikolai Sergeev and Guenther Palm
Institute of Neural Information Processing, Ulm University, Ulm, Germany

Keywords: Object recognition system, Invariant object representation.

Abstract: This paper presents a new 2D object recognition system. The object representation used by the system is rotation, translation, scaling and reflection invariant. The system is highly robust to partial occlusion, deformation and perspective change. The last makes it applicable to 3D tasks. Color information can be ignored as well as combined with form representation. The boundary of an object to be recognized doesn't need to be path-connected. The time demand to learn a new object doesn't depend on the number of objects already learned. No object segmentation prior to recognition is needed. To evaluate the system the 3D object library COIL-100 was used.

1 INTRODUCTION

1.1 System Architecture

An object recognition system usually consists of three parts. Part one extracts image primitives e.g. edges(Canny, 1986), lines(Hough, 1962), orientation histograms (Dalal and Triggs, 2005) or moments (Hu, 1962),(Reiss, 1993). Part two constructs feature vectors. Finally part three is responsible for learning and retrieval of the information e.g. support vector machines (Vapnik, 1998), artificial neural networks (Rosenblatt, 1962; Bishop, 2007) or regression estimators (Gyofri et al., 2002). The system to introduce in this paper also has this architecture. Image primitives are half ellipses. One feature vector encodes a combination of half ellipses. To learn and compare the feature vectors a new storage as well as a new retrieval algorithm were developed.

1.2 Motivation

Affine invariant object representation methods normally used are either not suitable for not path-connected objects as fourier descriptors (Arbter et al., 1990) or need segmentation prior to recognition as moments (Reiss, 1993). One common problem of these approaches is discrimination. Invariant features deliver no unique description of an object. So it can happen that two objects with similar features have nothing in common for an observer. The representation to be introduced in this paper overcomes theses problems. However it is not suitable for standard machine learning algorithms as support vector machines or neural networks. An image to analyze doesn't produce just a single representation vector but e.g. more than 50^{10} feature vectors. It makes standard retrieval algorithms unusable. For that reason a new type of storage together with a new search algorithm was developed.

In sum all the three characteristic components of this object recognition system (features, representation, machine learning algorithm) are new.

2 OUTLINE OF IMPLEMENTATION

An object is represented as a set of half ellipse combinations A as shown in Figure1. Combinations don't need to be of equal length.

For each $a \in A$ the system looks for a corresponding half ellipses combination b in the image to analyze. b should be as long as possible.

More precisely expressed: From the image to analyze the system extracts a set of half ellipses B as shown in Figure 2. For each combination $(a_i)_{i\in\{1,...,n\}} \in A$ a maximal $m \in \{1,...,n\}$ has to be determined for which a subsequence $\pi \in \{1,...,n\}^{\{1,...,m\}}$ with $\pi(1) = 1$ and $(b_i)_{i\in\{1,...,m\}} \in B^m$ exist so that $(a_{\pi(i)})_{i\in\{1,...,m\}}$ can be approximately transformed into $(b_i)_{i\in\{1,...,m\}}$ through translation, rotation, scaling, reflection and perspective change as

Sergeev N. and Palm G..
A NEW OBJECT RECOGNITION SYSTEM.
DOI: 10.5220/0003307103950400
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2011), pages 395-400
ISBN: 978-989-8425-47-8

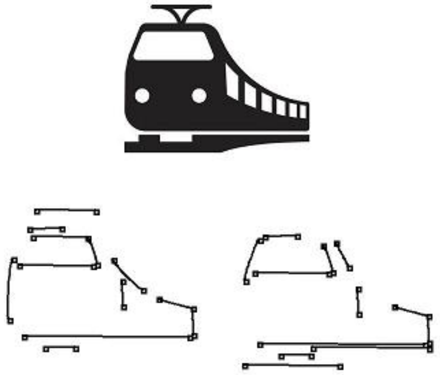

Figure 1: An object to learn and its representation.

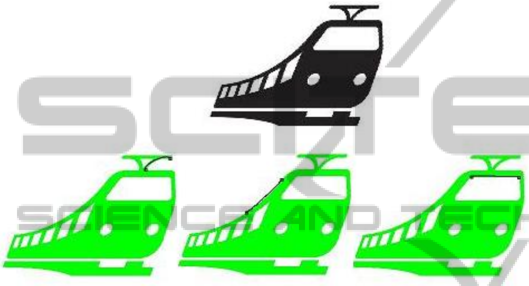

Figure 2: An object to analyze with a set of extracted half ellipses.

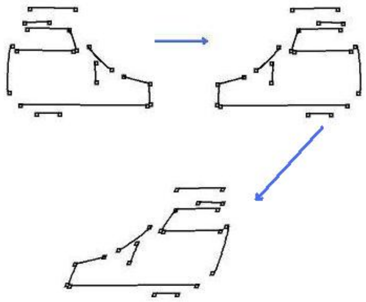

Figure 3: A way to transform one combination into anther.

shown in Figure 3. All in all the entire number of combination pairs to be compared is

$$\sum_{(a_1,\ldots,a_n)\in A}\left(\sum_{m=1}^{n}\binom{n-1}{m-1}|B|^m\right). \tag{1}$$

With A containing only one combination of length $n = 10$ and B consisting of 50 half ellipses the number of pairs is at least 50^{10}.

As the system makes the check for each subsequence $(a_{\pi(i)})_{i\in\{1,\ldots,m\}}$ it is robust to partial occlusion.

Without any further extension this representation is color invariant.

3 INVARIANT REPRESENTATION OF A COMBINATION OF HALF ELLIPSES

3.1 Overview

In this section a number of functions F^i are introduced. There are needed to obtain an invariant representation of a half ellipses combination. To understand their geometrical meaning it is unnecessary to read their mathematical description. The corresponding figures are enough. The most important figure is number 7. It shows the rotation, translation and scaling invariant representation of a combination of half ellipses.

3.2 Half Ellipses

For $\mathbb{C} = \mathbb{R}^2$ and $P(\mathbb{C})$ standing for power set of $\mathbb{C}$ a half ellipse is defined as a pair $(e, B) \in \mathbb{C}^2 \times P(\mathbb{C})$ with $e_1 \neq e_2$ for which $(a, b, t_0, \delta) \in [0,\infty) \times [0,\infty) \times \mathbb{R} \times \{-1, 1\}$ as well as $(c, \beta) \in \mathbb{C} \times \mathbb{R}$ exist so that

$$B = \left\{ T_c R_\beta \begin{pmatrix} a\cos t \\ b\sin t \end{pmatrix} \middle| \, t \in \begin{matrix} [t_0, t_0 + \delta\pi] \\ \cup \\ [t_0 + \delta\pi, t_0] \end{matrix} \right\} \tag{2}$$

and

$$e_1 = T_c R_\beta \begin{pmatrix} a\cos t_0 \\ b\sin t_0 \end{pmatrix}, \tag{3}$$

$$e_2 = T_c R_\beta \begin{pmatrix} a\cos(t_0 + \pi) \\ b\sin(t_0 + \pi) \end{pmatrix}. \tag{4}$$

T_c, R_β stand for translation, scaling and rotation respectively. The set of half ellipses will be denoted with HE. In other words a half ellipse consists of endpoints $e_1, e_2 \in \mathbb{C}$ and of a set of bow points $B \in P(\mathbb{C})$. The endpoints of a half ellipse play a very important role for the invariant representation. There are mainly two reasons to use half ellipses. An affine transformation $A \neq 0$ always maps a half ellipse onto another half ellipse. The second reason is the variety of half ellipses as Figure 4 shows.

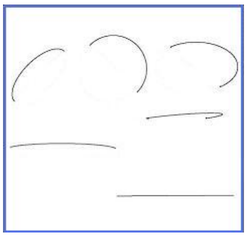

Figure 4: Examples of half ellipses.

3.3 Rotation, Translation and Scaling Invariant Representation of a Half Ellipse

Now a unique rotation, translation and scaling invariant representation of a half ellipse will be introduced. At first two preliminary definitions are needed. For $x,y \in \mathbb{C}$ with $x \neq y$ the function $F^1_{x,y}$ is defined as

$$F^1_{x,y} : \begin{cases} \mathbb{C} \rightarrow \mathbb{C} \\ z \mapsto \frac{z-x}{y-x} \end{cases} . \tag{5}$$

Figure 5 shows the geometric meaning of the transformation.

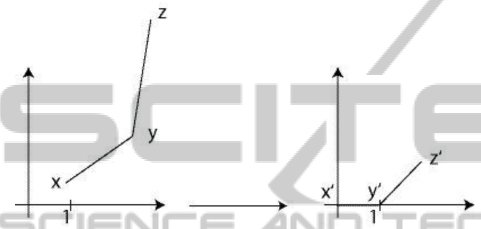

Figure 5: Geometric meaning of $F^1_{x,y}(z)$.

$F^1_{x,y}$ is an affine transformation with $F^1_{x,y}(x) = 0$ and $F^1_{x,y}(y) = 1$. The second function $F^2 : HE \rightarrow \mathbb{C}$ is defined as

$$F^2(e,B) = \begin{pmatrix} \max_{x\in B} \left| \left(F^1_{\frac{e_1+e_2}{2},\, e_1}(x) \right)_1 \right| \\ \\ \max_{x\in B} \left| \left(F^1_{\frac{e_1+e_2}{2},\, e_1}(x) \right)_2 \right| \end{pmatrix} . \tag{6}$$

Finally the invariant representation $F^3 : HE \rightarrow \mathbb{C}$ is defined in such a way that for always existent $x,y \in B$ with

$$F^2(e,B) = \begin{pmatrix} \left| \left(F^1_{\frac{e_1+e_2}{2},\, e_1}(x) \right)_1 \right| \\ \\ \left| \left(F^1_{\frac{e_1+e_2}{2},\, e_1}(y) \right)_2 \right| \end{pmatrix} \tag{7}$$

$$F^3(e,B) = \begin{pmatrix} z - SIGNUM(z) \\ \left(F^1_{\frac{e_1+e_2}{2},\, e_1}(y) \right)_2 \end{pmatrix} \tag{8}$$

with

$$z = \left(F^1_{\frac{e_1+e_2}{2},\, e_1}(x) \right)_1 . \tag{9}$$

For (e,B) in Figure 6

$$F^3(e,B) = \begin{pmatrix} M_1 - 1 \\ M_2 \end{pmatrix} . \tag{10}$$

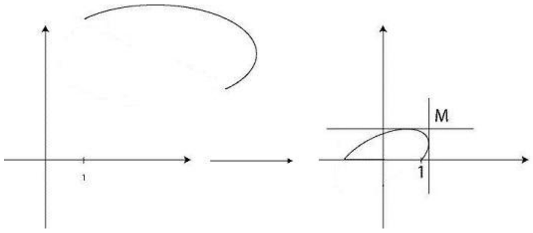

Figure 6: Representation of a half ellipse.

It can be shown that for each $x \in \mathbb{C}$ a half ellipse $(e,B) \in HE$ exists with $F^3(e,B) = x$. Additionally for two half ellipses $(e,B), (\tilde{e},\tilde{B}) \in HE$ with $F^3(e,B) = F^3(\tilde{e},\tilde{B})$ it can be shown that they can be transformed in each other through translation, rotation and scaling. On the other side $F^3(e,B)$ is invariant to translation, rotation and scaling of (e,B).

3.4 Extraction of a Half Ellipses

In the literature one can find several methods to extract ellipses. Most of them are based on Hough transform e.g. (Tsuji and Matsumoto, 1978). However they are not suitable for half ellipse detection as they do not deliver endpoints.

Endpoints of a half ellipse to be extracted don't need to be labeled or explicitly visible e.g. as corners. For that reason from a circle the system extracts several half circles. The exact number depends on the size of the circle. A bigger one can deliver over 100 half circles.

The invariant representation introduced above offers a convenient way to extract a half ellipse. As the Figure 6 shows the system determines two extremes for a chain of edge points. In the next step it calculates the unique half ellipse which would also have such extremes and endpoints. Then it checks if all the edge points of the chain are in an ε-neighborhood of the calculated unique half ellipse.

3.5 Rotation, Translation and Scaling Invariant Representation of a Combination of Half Ellipses

The set of combinations $C = \bigcup_{n\in\mathbb{N}} HE^n$ consists of ordered sequences of half ellipses. Rotation, translation and scaling invariant representation $F^4 : C \rightarrow \bigcup_{n\in\mathbb{N}} \mathbb{R}^{6n}$ is defined as

$$F^4\left((e^i,B^i)_i\right) = \left(F^1_{e^1_1,e^1_2}(e^i_1), F^1_{e^1_1,e^1_2}(e^i_2), F^3(e^i,B^i)\right)_i \tag{11}$$

with $i \in \{1,...,n\}$. Values of the representation of the

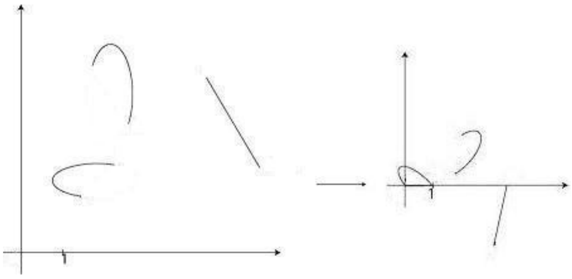

Figure 7: Representation of the endpoints of a combination.

endpoints of the combination showed in the left part of Figure 7 can be directly read off from the right part of the figure.

3.6 Reflection Invariance

The rotation, translation, scaling invariant representation introduced above will be now extended to a reflection invariant one. The purpose is to find a representation which doesn't change with the axis of reflection. For $n \in \mathbb{N}$ and diagonal matrix $M_n \in \mathbb{R}^{6n \times 6n}$ defined as

$$M_n = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & \dots & 0 & 0 \\ & & \ddots & & \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & 0 & -1 \end{pmatrix} \tag{12}$$

rotation, translation, scaling and reflection invariant representation F^5 is defined as

$$F^5 : \begin{cases} C \to \bigcup_{n \in \mathbb{N}} P(\mathbb{R}^{6n}) \\ c \mapsto \{F^4(c), M_n(F^4(c))\} \end{cases} . \tag{13}$$

In other words the representation consists of two feature vectors. On the one hand this representation doesn't change despite rotation,..., reflection. On the other hand two combinations with identical representations can be transformed in each other through rotation,..., reflection.

An example makes plausible why the additional vector is invariant to the axis of reflection. Figure 8 shows one combination reflected horizontally and vertically. Figure 9 shows the identical code of the endpoints of both reflected combinations.

4 ROBUSTNESS TO PERSPECTIVE CHANGE

Figure 10 offers a sketch of the rather technical formulation and implementation of the view point tolerance of the system. As shown in the left part of the figure the task is to recognize an object within the frame if a camera is placed at some point of the sphere above the dark line and its projection surface is parallel to the tangential plane of the point. To model a camera the perspective projection was used as it is described e.g. in (Jaehne, 2005).

To solve this task the system builds an in some sense minimal coverage of the sphere above the dark line as shown in the right part of the figure. For each point of the coverage the system makes a perspective transformation of the original rotation,..., reflection invariant representation and learns it. Albeit storage intensive this solution is simple and mathematically precise.

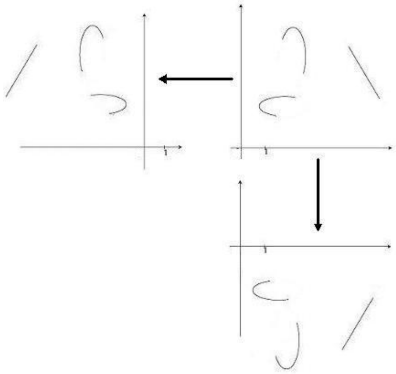

Figure 8: A combination reflected horizontally and vertically.

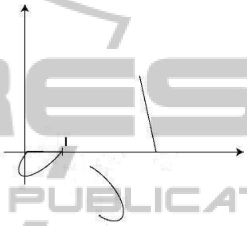

Figure 9: Representation of endpoints of the reflected combinations.

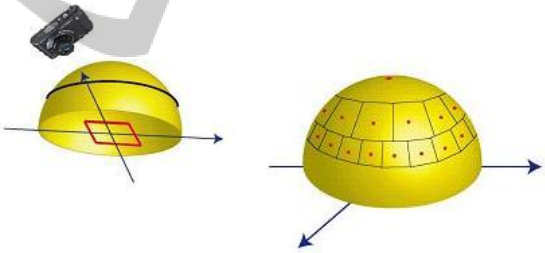

Figure 10: Formulation and implementation of perspective robustness.

5 THE TASK OF THE RETRIEVAL ALGORITHM

The storage of the system saves a set $A \subseteq \bigcup_{n \in \mathbb{N}} \mathbb{R}^{6n}$ of feature vectors representing combinations of half ellipses not the original combinations.

From an image to analyze the system extracts a set of half ellipses $B \subseteq HE$. For each $a = (a_i)_{i \in \{1,\dots,n\}} \in A \cap \mathbb{R}^{6n} = A \cap \prod_{i \in \{1,\dots,n\}} \mathbb{R}^6$ the retrieval algorithm determines maximal $m \in \{1,\dots,n\}$ for which a subsequence $\pi \in \{1,\dots,n\}^{\{1,\dots,m\}}$ with $\pi(1) = 1$ and $b \in B^m$ exists with

$$\forall i \in \{1,\dots,m\} : \left\| a_{\pi(i)} - c_i \right\|_{\max} \leq \varepsilon \qquad (14)$$

with $F^4(b) = c \in \mathbb{R}^{6m}$ and $\varepsilon > 0$. With other words two feature vectors get compared with respect to maximum norm.

To find such a maximal $m \in \{1,\dots,n\}$ the system tries all m, π, b. The new type of machine learning algorithm allows to compare each $\left(a_{\pi(i)}\right)_i$ with each $c = F^4(b)$. Thus the task is to find the highest m with a successful comparison.

$\varepsilon > 0$ can be chosen only once prior to the initialization of the system.

Comparing two feature vectors with respect to maximum norm the system tolerates deformation of an object within ε.

6 EXPERIMENTAL RESULTS

6.1 COIL-100

To evaluate the system the well known database COIL-100 (Columbia Object Image Library) was used. It is available at http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php. The data set is described in (Nene et al., 1996). It contains 7200 color images of 100 3D objects shown in Figure11. One image is taken per 5° of rotation.

6.2 Experiment Settings and Results

The computer used in the experiments has a processor Intel(R) Core(TM)2 Duo CPU P8600 @2.40 GHz 2.40 GHz and 4.00 GB RAM. The system is implemented in Java.

There were made 2 experiments with slightly different parameter settings.

In the first experiment 18 views(1 per 20°) were used to learn each object. The remaining 5400 images were analyzed. A recognition rate of 99.2% was reached. The time demand to learn all objects is 277 seconds. The average time demand to analyze one image is 980 milliseconds.

Figure 11: COIL-100 objects.

In the second experiment 8 views(1 per 45°) were used to learn an object. The other 6400 were analyzed. A recognition rate of 96.3% was reached. The system needs 142 seconds to learn all objects. The time demand to analyze a single image is 1593 milliseconds.

6.3 Comparison to other Methods

The Table 1 is based on the results described in (Yang et al., 2000) and (Caputo et al., 2000).

Table 1: Comparison with Alternative Results.

Method	18 views	8 views
LAFs	99.9%	99.4%
Half Ellipses	99.2%	96.3%
SNoW / edges	94.1%	89.2%
SNoW / intensity	92.3%	85.1%
Linear SVM	91.3%	84.8%
Spin-Glass MRF	96.8%	88.2%
Nearest Neighbor	87.5%	79.5%

6.4 Color Information

The pure form representation described above was extended with color information. A half ellipse has the first and the last point. Hence it also has the right and the left side as the Figure12 shows. After the extraction of a half ellipse the system determines arithmetic RGB average along the right side of the half ellipse as well as along the left one. Thus it determines two RGB vectors $l, r \in \mathbb{R}^3$. Color code $c \in \mathbb{R}^6$ is just Cartesian product of this two vectors $c = (l, r)$. A representation vector $a \in \mathbb{R}^{6n}$ of a half ellipse combination $b \in HE^n$ gets extended to $\tilde{a} \in \mathbb{R}^{6n+6n}$ with color code $(c_i)_{i \in \{1,\dots,n\}} \in \prod_{i \in \{1,\dots,n\}} \mathbb{R}^6$ for each half ellipse of the combination. An additional threshold

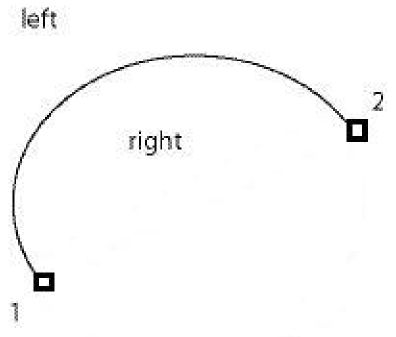

Figure 12: The left and the right side of a half ellipse.

value $\tilde{\varepsilon} > 0$ is used to compare the color information of two representation vectors with respect to maximum norm. An upcoming article explicitly describes the extraction of half ellipses and its color.

6.5 Learning and Recognition Scheme used for COIL-100

As mentioned above the system uses e.g. 8 images to learn an object. For one image it constructs e.g. 10 combinations of half ellipses. Each combination is represented with e.g 6 feature vectors. Each vector is labeled with the number $N \in \{1, ..., 100\}$ of the object it refers to.

Analyzing an image the system at first determines the maximal length $m \in \mathbb{N}$ of the matched subsequences for each learned feature vector. Let the set of such lengths be denoted as M. For $\tilde{m} = \max M$ the system depicts all feature vectors for which subsequences of the length $\tilde{m}$ were matched. The object with the greatest number of such feature vectors will be returned as the recognized one. Having several such objects the system chooses one of them randomly.

7 SUMMARY AND FUTURE WORK

The object recognition system presented in this paper combines several important characteristics. It's capable of handling 3D objects. The half ellipse extraction is at least stable enough to wield COIL-100 images.

The trivial color representation used now has yet to become illumination invariant. The optimization of the running time doesn't appear to be a great problem as the central retrieval algorithm is highly parallelizable. The greatest challenge seems to be the reduction of the storage consumption without lost of perspective robustness.

At the present the authors develop a flow estimator based on the comparison of half ellipse combinations. The flow estimator learns thousands of half ellipse combinations on the first frame and tries to match them on the second one. So in a near future the system could gain an universal character being simultaneously capable of object recognition as well as flow estimation.

REFERENCES

Arbter, K., Snyder, W. E., Burkhardt, H., and Hirzinger, G. (1990). Application of affine-invariant fourier descriptors to recognition of 3d objects. In *IEEE Trans. Pattern Analysis and Machine Learning*.

Bishop, C. (2007). *Neural Networks for Patternrecognition*. Oxford University Press.

Canny, J. (1986). A computational approach to edge detection. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Caputo, B., Hornegger, J., Paulus, D., and Niemann, H. (2000). A spin-glass markov random field for 3d object recognition. *NIPS 2000*.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Conference Computer Vision and Pattern Recognition , San Diego*.

Gyofri, L., Kohler, M., Krzyzak, A., and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer.

Hough, P. V. C. (1962). *Method and Means of Recognising Complex Patterns*. US Patent 3069654.

Hu, M. K. (1962). Visual pattern recognition by moment invariants. In *IRE Transactions on Information Theory*.

Jaehne, B. (2005). *Digital Image Processing*. Springer-Verlag Berlin.

Nene, S. A., Nayar, S. K., and Murase, H. (1996). *Columbia Object Image Library (COIL-100)*.

Reiss, T. H. (1993). *Recognizing Planar Objects Using Invariant Image Features*. Springer-Verlag Berlin Heidelberg.

Rosenblatt, F. (1962). *Principles of Neurodynamics*. Spartan, New York.

Tsuji, S. and Matsumoto, F. (1978). Detection of ellipses by a modified hough transform. In *IEEE Trans. Comput.*

Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley, New York.

Yang, M. H., Roth, D., and Ahuja, N. (2000). Learning to recognize 3d objects with snow. In *ECCV 2000*.