第21卷第5期　　红外与毫米波学报　　Vol. 21, No. 5
2002年10月　　J. Infrared Millim. Waves　　October, 2002

# A SIMILARITY MEASUREMENT FOR AFFINE INVARIANCE RETRIEVAL BASED ON SYNERGETIC NEURAL NETS(SNN) *

QI Fei-Hu[1)]　ZHAO Tong[1)]　Horace H S Ip[2)]
([1)] Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, 200030, China;
[2)] Image Computing Group, Department of Computer Science, City University of Hong Kong, Hong Kong)

Abstract　A new concept of control parameter, order vector, based on SNN was proposed. More specifically, its associate properties for image retrieval was studied. Based on the properties, an efficient element-based affine invariance similarity measurement was proposed for retrieval of trademark images. The experiments indicate that the retrieval method is good in against noisy and occluded images, and the method has affine invariance in translated, rotated and scaled images.
Key words　Image retrieval, Similarity measurement, Affine invariance, synergetic neural net (SNN).

# 基于协同神经网络的仿射不变性检索相似性度量方法 *

戚飞虎[1)]　赵　同[1)]　Horace H S Ip[2)]
([1)]上海交通大学计算机科学与工程系,200030;
[2)] Image Computing Group, Department of Computer Science, City University of Hong Kong, Hong Kong)

摘要　根据协同神经网络,提出了一种新的控制参数——序矢量,并着重研究了其在图像检索方面的相关性质.在此基础上,针对商标图像检索,提出了一种高效的基于元素的仿射不变性度量方法.实验表明,该检索算法抗噪、抗缺损能力强,同时对于平移、旋转、缩放具有不变性.
关键词　图像检索,相似性度量,仿射不变性,协同神经网络.

## Introduction

A visual keywords-driven image retrieval approach has been proposed in our previous works[1,2]. The approach has many advantages such as robustness in against noise and occlusion affine invariant, as well as good scalability. To yield good performance, one of the key issues of our approach on retrieval is the design of a suitable similarity function within the theoretical framework of synergetic neural nets (SNN). In this paper we develop a formal foundation which allows us to construct a similarity function for SNN-based image retrieval using the concept of visual keywords.

## 1　Control parameters in SNN

### 1.1　SNN

Synergetics is an interdisciplinary field of research that is concerned with the spontaneous formation of macroscopic spatial, temporal, or functional structures of systems via self-organization[3]. Haken suggests[4] that there are large classes of system in which self-organization obeys the same basic principles. For a test pattern q, it can be constructed by a dynamics, which pulls the test pattern via intermediate state q(t) into one of the prototype patterns $\nu_k$:

$$E = -\frac{1}{2}\sum_{k=1}^{p}\lambda_k(v_k^+ \cdot q)^2 + \frac{1}{4}B\sum_{k\neq k'}(v_k^+ \cdot q)^2(v_{k'}^+ \cdot q)^2 + \frac{1}{4}C(q^+ q)^2, \quad (1)$$

where E is potential function, p is the number of the prototypes, B and C are constants and $\lambda = [\lambda_1, \lambda_2, \cdots, \lambda_p]$ is a vector of constants labeled attention parameters.

*国家自然科学基金(批准号69772002)资助项目
稿件收到日期2002-02-26,修改稿收到日期2002-03-20

*The project Supported by National Natural Science Fund of China (No.69772002)
Received 2002-02-26, revised 2002-03-20

1.2 Adjoint vectors

In Equation (1), $v_k^+$ is called adjoint vector, which obeys the orthonomality relations:

$$(v_k^+ \cdot v_i) = \delta_{ki}, \tag{2}$$

where $\delta_{ki}$ is the Kronecker delta function.

Lemma 1. The matrix $V^+ = [(v_1^+)'(v_2^+)' \cdots (v_p^+)']'$ is the generalized inverse matrix to the matrix $V = [v_1 v_2 \cdots v_p]$ if $v_1, v_2, \cdots, v_p$ is linearly independent.

1.3 Order parameters

When a system is controlled by a set of control parameters, at specific values of those parameters, the system may undergo qualitative macroscopic changes. The old state loses its stability, and in the transition, the dynamics are governed by a few collective variables: the order parameters

$$\xi_k = v_k^+ q. \tag{3}$$

The $\xi_k$'s enslave all the other variables of the system, so that an enormous reduction of the degrees of freedom is obtained.

## 2 Extraction of affine invariance features

In our approach, affine invariant features in the Fourier domain is first extracted. Concretely, if $f_1(x, y)$ is a affine transformed replica of $f(x, y)$ with translation$(x_0, y_0)$, rotation $\theta_0$ and an uniform scale factor $a_0$, then

$$f_1(x, y) = f\left(\frac{x\cos\theta_0 - y\sin\theta_0}{a_0} - x_0, \frac{x\sin\theta_0 + y\cos\theta_0}{a_0} - y_0\right).$$

According to the affine properties of Fourier transform, transforms of $f_1(x, y)$ and $f(x, y)$ are related by

$$F_1(u, v) = \frac{1}{a_0^2} e^{-j2\pi(ux_0 + vy_0)} \times F\left(\frac{u\cos\theta_0 - v\sin\theta_0}{a_0}, \frac{u\sin\theta_0 + v\cos\theta_0}{a_0}\right).$$

By the normalizing magnitudes of F、$F_1$, mapping Euclidean coordinates to polar coordinates via the logarithmic function and converting the axes to logarithmic scale, $F(u, v)$ and $F_1(u, v)$ is transformed into $M(r, \theta)$ and $M_1(r, \theta)$,

$$M_1(r, \theta) = M(\log r - \log a_0, \theta - \theta_0),$$

the rotation and scale can be reduced to translation in complex logarithm polar coordinates. Thus, affine invariance can be obtained by using Fourier transforms in such space.

## 3 Order vectors

Several synergetic models for pattern recognition have been reported in the past, such as deformed patterns of handwriting characters[5], pose estimation of 3D object[6], voice recognition[7]. All of these models make use of the order parameter value as the control parameter. A major challenge in content-based retrieval for trademark images is that we look for images that are perceptually similar rather than exactly the same as required in pattern recognition applications. This required a better understanding of the interpretation of the order parameter for perceptual similarity.

3.1 Concept

First, we would like to introduce a concept of dot-vector product:

Definition: we assume vector $a = (a_1, a_2, \cdots, a_n)$ and $b = (b_1, b_2, \cdots, b_n)$, then the dot-vector product of a and b is defined as

$$O(a, b) = a \cdot b = (a_1, a_2, \cdots, a_n) \cdot (b_1, b_2, \cdots, b_n) = (a_1 b_1, a_2 b_2, \cdots, a_n b_n),$$

Thus, we introduce a new control parameter, more specifically control vector, called Order Vector:

$$O(v_k^+, q) = v_k^+ \cdot v = (v_{k1}^+ q_1, v_{k2}^+ q_2, \cdots, v_{kn}^+ q_n).$$

3.2 Properties

Lemma 2 All the order vector $O(v_k^+, q)$ has the same affine invariant features distribution, that is, the same ith element $o_i(v_k^+, q_{i1})$ and $o_i(v_k^+, q_{i2})$ represent the same affine invariant feature in different order vector $O(v_k^+, q_{i1})$ and $O(v_k^+, q_{i2})$, respectively.

Proof As the affine invariant features are based on Fourier transformation and complex logarithm mapping, $v_{ij}$, the jth element of $v_i$, will represent the value of the jth affine invariant feature in $v_i$, $F_j^{inv}|_{v_i}$.

From Lemma 1, the adjoint vectors $v_k^+$ can be represented as linear superposition of the transposed visual keywords $v_k'$: $v_k^+ = \sum_j a_{kj} v_j'$, that is $v_k^+$ is the linear combinations of visual keywords, thus $v_{ki}$ implies the properties of $F_i^{inv}|_{v_k^+}$, which also interprets that $v_k^+$ has the same element representation as $v_k$.

As $O(v_k^+, q) = v_k^+ \cdot q$, $o_i(v_k^+, q_{i2})$ will deep the ith element properties of $v_k^+$ and q, and it represents $F_i^{inv}|_{O(v_k^+, q_{i1})}$. As the transformation remain the same to any order vector, $o_i(v_k^+, q_{i1})$ and $o_i(v_k^+, q_{i2})$ will represent the different values of the same features in different order vectors, that is, $F_i^{inv}|_{O(v_k^+, q_{i1})}$ and $F_i^{inv}|_{O(v_k^+, q_{i2})}$, respectively.

Theorem 1 Each order vector $O(v_k^+, q)$ consists of two part: one is its positive part $O^{pos}$ and the other part is the negative part $O^{neg}$, where $O^{pos}$, $O^{neg}$ represent the similar and dissimilar part of the two images, $v_k$ and q respectively.

Proof The SNN considered the prototypes as a whole[2]. $v_k^+$ extracts the significant features of $v_k$ apart from other prototypes, that is, the similarity degree of $v_k$ and q is replaced by that of $v_k^+$ and q. For a given query q, as $v_k^+ = \sum_j a_{kj} v'_j$, the order vector can be written as

$$O(v_k^+, q) = v_k^+ \cdot q = a_{kk} v'_k \cdot q + \sum_{j \neq k} a_{kj} v'_j \cdot q. \quad (4)$$

If in the ith element, $q_i$ has the same sign with $v_{ki}$, from Lemma 2 and $a_{kk} > 0$[2], we can know in the ith affine invariance feature, $q_i$ has the same properties with $v_{ki}$, which may cause the first term of equation (4) positive. Moreover, if $O_i > 0$, that means $a_{kk} v'_{ki} \times q_i > -\sum_{i \neq k} a_{ki} v'_{ii} \times q_i$, which implies the similarity degree in ith feature between q and $v_k$ will be greater than that between q and all other prototypes properties $v_i$. Otherwise if $o_i < 0$, the ith property of q is dissimilar with $v_k$.

Therefore, we can divide order vector $O(v_k^+, q)$ into two parts: one is its positive part $O^{pos} = \{O_i | O_i > 0\}$ and the other part is the negative part $O^{neg} = \{O_i | O_i < 0\}$, which represents the similar and dissimilar part of the two images, respectively.

## 4 Construction of Similarity Measure

### 4.1 Similarity measurement

The aim of image retrieval as compared to image recognition is that image retrieval systems interest not only on total but also partial shape similarity of the images in the database. In such cases, one should pay more attentions on the similarities than on the differences. Therefore, similarity measurement should mainly reflect the degree of likeness between two images.

Based on Theorem 1, we can know the positive part $O^{pos}$ of the order vector describes the similarity of two images, which is just the interested part when human judge the similarity of two images. Thus, for a given query q, we construct a similarity measurement $S(v_k, q)$ for each prototype $v_k$ by only using $O^{pos}$:

$$S(v_k, q) = \sum_i O_i(v_k, q). \quad O_i > 0$$

### 4.2 Similarity threshold

From Theorem 1, the elements of the order vector are associated with the affine invariance features. In fact, except the similar and dissimilar parts, there still exists another part: quasi-zeros part $O^{zero} = \{O_i | O_i < \delta\}$, where $\delta$ is a small positive constant. Some of $O^{zero}$ are caused by the quantization error, and the others are caused by small value of $F_i^{inv}$ in either $v_k$ or q. For both cases we would neglect such small values which contribute insignificantly or even donot harm to the perception of similarity, so we redefine more useful order vector with a threshold $\delta$ as:

$$O^{\delta}(a, b) = \{O_i^{\delta}(a, b), i = 1, \cdots, m\},$$

$$\text{where } O_i^{\delta}(a, b) = \begin{cases} a_i b_i & \text{if } a_i b_i > \delta \\ 0, & \text{otherwise} \end{cases}$$

Thus we can define the similarity function based only on the significant similar components within the order vector:

$$S^{\delta}(v_k, q) = \sum_i O_i^{\delta}(v_k^+, q) \quad (5)$$

### 4.3 Normalization

In image retrieval, similarity ranking is an important factor to help determining the degree of similarity among the retrieved patterns with all the images in database $D^{img}$. Based on the similarity measure, we can know the similar degree between $v_k$ and all the image $q_i$ in $D^{img}$. However, the similarity vector $S^{\delta}(v_k, q) = \{S^{\delta}(v_k, q), k = 1, \cdots, p\}$ implies the percentage of the image q similar with the each visual keywork $v_k$, $k = 1, \cdots, p$. If $S^{\delta}(v_1, q) > S^{\delta}(v_2, q)$, it can be sure that q is much similar with $\nu_1$ than $\nu_2$, and vice visa. But if $S^{\delta}(\nu_1, q_1) > S^{\delta}(\nu_1, q_2)$, as $\sum_{k=1}^{p} S^{\delta}(\nu_k, q_1) \neq \sum_{k=1}^{p} S^{\delta}(\nu_k, q_2)$, it is difficult for us to tell whether query $q_1$ or $q_2$ is more like $\nu_1$. In order to compare those similarity values, it is necessary to normalize all the $S^{\delta}(\nu_k, q)$ on a same level:

$$SN^{\delta}(\nu_k, q) = \frac{S^{\delta}(\nu_k, q)}{\sum_{k=1}^{p} S^{\delta}(\nu_k, q)} . \tag{6}$$

## 5 Experimental results

### 5.1 Database

To test the performance of the novel similarity measurement, we investigated a database with 131 trademark images. Each image is 64 ×64 size and 256 gray levels

### 5.2 Visual keywords-based retrieval system

Image retrieval often deals with a large-scale database. Neural network applied for image retrieval has its advantages in against noisy and occluded images, but the size of neural network may be increased greatly as the new prototypes add. Ordinarily, geometric graphs, such as circle, square, triangle, and so on (shown in Fig. 1), exist in most trademark images, and we introduced such geometric graphs as the visual keywords for image retrieval. From our previous experiments, SNN can recognize partial affine images, so we can extend this capability to using visual keywords as prototypes to search for trademark images which consist of such visual keywords[1,2].

### 5.3 Bull retrieval by human

As the similarity rank is subject to human perception, here we adopted the standard test patterns and retrieval results from A. K. Jain's work[8]. The trademarks shown in Fig. 2 are topmost retrieved trademark images by five human subjects for the bull template. We combined the 10 images shown in Fig. 2 into our existing trademark image database, which may give us more convictive experiment results.

Table 1 20 topmost similar trademark images corresponding to visual keywords

表 1 与视觉关键词最相似的 20 个商标图


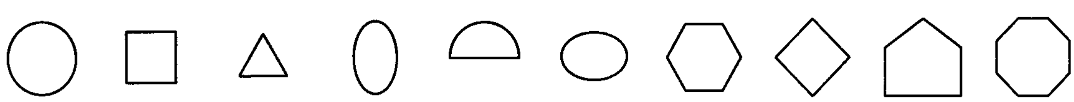

Fig. 1 Visual Keywords

图 1 视觉关键词

Fig. 2 Topmost retrieval trademark images by human subjects
图2 测试者检索率最高的商标图

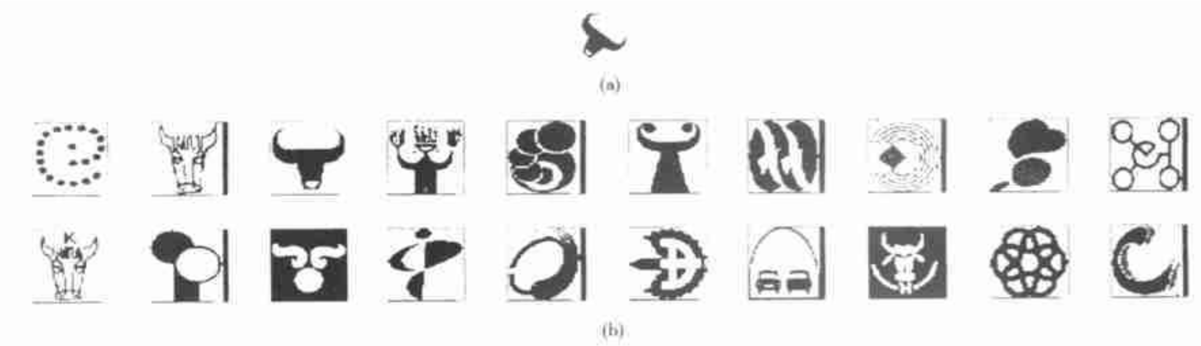


Fig. 3 Database retrieval results (a) an affine transformed bull image , (b) top-20 retrieval results
图3 数据库检索结果 (a) 经过仿射变换的公牛图 (b) 列前20位检索结果

5.4 Retrieval based on keywords

Here we would like to choose any image of Fig. 2 , square and triangle as the visual keywork prototypes(seen in the column 1 of Table 1) to train the SNN , and then evaluate the similarity value of each image in database upon the proposed element-based similarity measure.

5.5 Affine invariance retrieval

To test our mentioned affine invariance properties, we translated , scaled and rotated any one of these ten images as one of the visual keywork images to test the retrieval property of the proposed method.

In the case of the query on a bull image (Fig. 3a) , the system extracts all the other images of a bullhead found in the database. These are the bull head images retrieved in the $2^{nd}$ ,$3^{nd}$ ,$4^{th}$ ,$6^{th}$ ,$11^{th}$ ,$13^{th}$ ,and $18^{th}$ places. Additionally , except for the bullhead images , almost all other retrieved images are also partly like affine transformed version of the bull 's horn structure in the query image. More retrieval results will be made available in the final paper.

## 6 Conclusions

The proposal highlight is the development of an order vector based similarity measure for image retrieval using a synergetic neural net. When it is used in conjunction with a set of affine invariance features , and embedded within the elements of the order vector , affine invariance retrieval of images can be achieved. The approach gives good retrieval performance and ignores the unexpected influence caused by dissimilar and irrelevant features , and noise. Further experimental results will be shown in the final paper.

## REFERENCES

[1] ZHAO Tong , Horace H S Ip , QI Fei-Hu. Synergetic neural network approach for content-based retrieval of trademarks. Invited paper. In : Proceedings of the Fifth Joint Conference on Information Sciences ,Atlantic City , U. S ,2000 ,2 :484.

[2] ZHAO Tong , TANG Lilian H , Horace H S Ip , et al. Visual keyword image retrieval based on synergetic neural network for web-based image search. Real-Time System , 2001 ,21 :127 —142

[3] Haken H. Synergetics : An Introduction. Berlin Heidelberg : Springer-Verlag ,1983

[4] Haken H. Synergetic Computers and Cognition —A Top-Down Approach to Neural Nets. Berlin Heidelberg: Springer-Verlag , 1991

[5] Daffertshofer , Haken H. A new approach to recognition of deformed patterns. Pattern Recognition , 1994 , 27 (12) : 1697 —1705

[6] Hogg T , TalhamiH , Ress D. An improved synergetic algorithm for image classification. Pattern Recognition , 1998 , 31(12) :1893 —1903

[7] Ditzinger T , Tuller B , Haken H , et al. A synergetic model for the verbal transformation effect. Biological Cybernetics , 1997 ,77 (1) :31 —40

[8] Jain Anil K , Vailaya A. Shape-based retrieval : a case study with trademark image databases. Pattern Recognition , 1999 ,31 (9) :1369 —1390