Integrating Visual Ontologies and Wavelets for Image Content Retrieval

Weng Zu Mao and David Bell
School of Information and Software Engineering
University of Ulster at Jordanstown
UK BT37 0QB
{zm.weng, da.bell}@ulst.ac.uk

Abstract

One of the challenging problems for image content retrieval is how to represent semantics of images. We have proposed a new representation for image contents using visual ontologies and wavelets. We will discuss how to represent the shape of objects/regions of interest, colour distribution and texture of the images integrating visual ontologies with wavelets. Our ideas are: (1) propose a set of meta ontologies and semantic descriptors (graphic and symbolic) for visual ontologies; (2) using visual ontologies and most significant coefficients of a 2-D multiresolution wavelet transform to represent image contents. The visual ontologies are represented by a set of descriptors, relations, logic operators, and functions; (3) using a multiresolution wavelet transform to extract the colour and texture features of target images and map it into relevant visual ontologies. The query images can also be represented as visual ontologies or most significant coefficients of a 2-D multiresolution wavelet; (4) a spatial query can be integrated with other visual ontologies. The initial experiments have shown encouraging results.

1. Introduction

How to represent the knowledge of image contents more effectively is a difficult problem that has been investigated for several years. Many literatures have been focused on the low-level features of images. The basic concepts are motivated by the results of image processing and computer vision. The individual definitions for images' semantic contents lead to the gaps among different systems. The demands for overcoming the difficulties of eliminating different understandings of image contents have motivated us to investigate a shared knowledge representation. On the other hand, interoperability among systems, reusability of software components, and reliability of systems have become an important evaluating criteria for distributed content-based image retrieval (CBIR) systems.

From a point of view for human-computer interaction, low-level features of images should be transformed into visual representation of knowledge. A set of well-defined semantic descriptors and visual ontologies (VOs) are required in a mapping process. These descriptors serve as a bridge between low-level representation and a visual user interface.

An increasing interest on ontology and its applications, a unifying framework for reducing or eliminating conceptual and terminological confusion towards shared understanding of knowledge, has motivated the investigation of VOs for image contents. Compared to symbolised ontology, VOs will be able to represent complex image knowledge in a more detailed and intuitive way. There are no image processing experts involved to process a complicated knowledge representation of images. A shared understanding of images could be achieved in this way.

Wavelets have recently become enormously popular. The name "wavelets" was introduced in the early eighties by Morlet. Grossman has provided the mathematical basis of Morlet's idea [11]. Daubechies constructed families of orthonormal wavelets with compact support [12]. Mallat put the wavelet transform (WT) in the framework of multiresolution signal decomposition and provided a fast implementation of the transform based on the concepts of filtering theory, which was already well-known in the image compression literature [15]. All this research resulted in a comprehensive mathematical framework for signal analysis with applications in variant fields.

In this paper, we mainly discuss the VOs for the shapes of objects/regions of interest, colour distribution, and texture of images; the 2-D multiresolution Haar wavelet decomposition for colour and texture representations; and the mapping from visual ontology based space to wavelet-based coefficients. In order to define the VOs, additional foundational descriptors are required to provide semantics of these ontologies. In a distributed CBIR system based on a software agent, such ontologies are required to enable the unambiguous communication between agents.

Many approaches for CBIR have been proposed and a couple of prototypes have been investigated as well. There are several representative prototypes among these prototypes: IBM's QBIC project [17], MIT's Photobook system [18], the Chabot system developed by UC Berkeley [16], and VisualSEEk and WebSEEk system developed by Columbia University [13]. Jacob et al. presented a multiresolution Haar wavelet decomposition method to speed up colour distribution matching [10].

Our integrated VOs and the multiresolution wavelet transform (MWT) approach emphasise the visual representation of image contents and fast multiresolution querying. Our approach also allows users to select spatial, shape, colour, and texture ontologies for composite terms or single term querying. The MWT approach has also been used for colour and texture retrievals.

This paper is organised as follows: section 2 explains what the VOs for CBIR are. In section 3, a set of semantic descriptors for VOs has been proposed. In sections 4, 5 and 6, we will present the VOs for the shapes of objects/regions of interest, colour distribution and texture of images, respectively. The approach of the 2-D MWT has been integrated with our VOs. Section 7 contains the result of initial experiments for integrating VOs and MWT. Finally, section 8 ends with the conclusion.

2. The VOs for CBIR

The semantic contents of images are initially represented by a set of low-level image features for CBIR so far. The features used most frequently include the shape, colour histogram, and texture of images. For the purpose of extraction of these low-level features, several dozens of image processing techniques have been developed. The domain experts are usually required be involved in the complicated knowledge representation of images. At the same time, rich image contents can not be well represented by a limited number of key words/captions. Users find it difficult to extract these semantic contents or select suitable features to describe their demands for retrieval. A symbolic and axiom-based representation of spatial systems has been proposed using ontology-based knowledge representations [1]. The spatial reasoning has been investigated in some literatures [2, 3, 4]. These literatures focused on the spatial-related and reasoning-based solutions. These solutions have been the basis of spatial representation of our VOs.

2.1. Meta ontologies for VOs

The visual ontology we propose here is an ad hoc paradigm of knowledge representation for CBIR. The ontologies of objects/regions of interest in an image can be considered as a set of basic visual elements — meta objects. These meta objects consist of a set of entities and can be partially described as follows:

Scalable objects with regular shapes: circle, rectangle, triangle, ellipse, line, arc, angle, etc.

Colours: red, green, blue components of any colours; hue, chroma, and saturation of any colours.

Rough colours: red, orange, yellow, green, blue, purple, pink, brown, gray, gold, black, white, etc.

Textures: natural or artificial textures such as clothes, land, cornfield, and wallpapers.

The meta objects can be considered as a set of meta ontologies. The meta ontologies also consist of all operators, complex numbers, and possible constraints.

The relationships among meta objects could be classified into:

Rough location relations: connects, overlaps, separates, contains, besides, over, below, left, right.

Precise location relations: relative or absolute coordinates of objects, angle between two objects .

Rough size relations: larger than, smaller than.

Rough colour relations: light, dark, more light than, more dark than, and, or, not.

Texture relations: foreground, background, and, or, not.

2.2. Functions and relations for VOs

Functions and relations for the VOs are all inter-relationships among objects. Because in the VOs based image retrieval system, an object is approximately assembled by a set of meta objects. The functions actually represent the interrelationships among meta objects. A function is defined as a limited set of objects and computation formulas, one for each combination of possible meta objects. The relations are an arbitrary set of limited lists of objects, each list is a selection of meta objects that jointly satisfy the relation.

3. Semantic descriptors for VOs

We define the *VOs* with a set of semantic descriptors that can be unambiguously interpreted. The descriptors for VOs can be classified into two categories: graphic and symbolic. The graphic descriptors are a set of graphic meta objects including shape, colour, texture, and spatial descriptors. Examples of shape graphic descriptors can be depicted by the basic graphic elements, as shown in Figure 1.

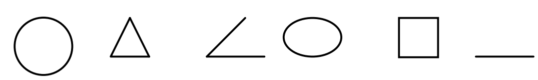

Figure 1. Basic graphic elements

All VOs for shape retrieval can be approximately assembled and depicted by these graphic descriptors. Any shape of object in images could be approximately segmented into these descriptors as well. Texture descriptors include the texture library with large collections of natural or artificial texture elements. Various colour descriptors could be customised with colour elements or continuous adjustable RGB components.

Another category of descriptors is symbolised. These descriptors describe the meta objects and their relations.

A graphic descriptor can be mapped into the corresponding symbolic descriptor.

Table 1. Description for meta objects

meta object	colour	texture	location relations	texture relations	colour relations
circle	*red*	*cloth wood*	*overlap*	*and*	*light*
trian-gle	*green*	*forest field*	*centre*	*or*	*dark*
rectan -gle	*blue, yellow*	*ocean, sand*	x_1, y_1	*and*	*and*

If we define a set of descriptors for meta objects, Table 1 above can be further symbolised into a set of simple descriptors, as shown in Table 2.

Table 2. Symbolic representation for VOs

meta object	Colour	texture	location relations	Texture relations	colour relations
M_{cir}	C_{red}	T_{cloth} T_{wood}	$L_{overlap}$	$\wedge$	CR_{light}
M_{tri}	C_{green}	T_{forest} T_{field}	L_{centre}	$\vee$	CR_{dark}
M_{rec}	C_{blue}, C_{yellow}	T_{ocean}, T_{sand}	x_1,y_1	$\wedge$	$\wedge$

Obviously, the representation for objects/region of interest and their relationships in an image become more simplified. Although we can not depend on these simple symbols to describe rich contents of images, it should facilitate an unambiguous, shared understanding of images. A set of well-defined descriptors should be available to represent semantic contents of an image more efficiently.

4. The Shape Ontologies (SOs)

The shape of an object/region of interest can be assembled approximately by a set of meta objects with corresponding relationships. Visual SOs represent these meta objects and their relationships more easily than other approaches.

The descriptors and relations can be used to match with indexed descriptors corresponding to the target images in an image database.

4.1 Definition of visual SOs

We define the visual SOs with a group of visual specifications, relationships, and functions. Users will be able to assemble meta ontologies interactively according to their desired shape.

Any complicated shape of objects could be similarly defined in this way. The location relationships among individual assembled meta ontologies will be tracked and located by their relative (or absolute) coordinates (or centroids), precisely. Rough locations relationship among objects can also be represented by the corresponding descriptors. The functions of SOs define a set of objects and computation formulae for the corresponding ontology.

Table 3. Definitions of SOs

Descriptor	Definition	Symbol	Location relation	Function
edge	*Boundary of object or region of interest*	S_e	*Lr set*	F_e
object	*Discri-minatable entity of interest in a region*	S_o	*Lr set*	F_o
region	*Area consisting of one or more objects of interest*	S_r	*Lr set*	F_r

Lr set: A set of rough location relations between objects or regions of interest such as left, right, above, contains, below, behind, overlap, separate, in front of, intersected.

In Table 3, functions F_e, F_o, F_r are a set of formulas as follows:

$$F_e = \sqrt{(X_s - X_c)^2 + (Y_s - Y_c)^2}$$

Where X_s, Y_s are sample coordinates of the shape (edge). X_c, Y_c are centre coordinates of the shape (edge). The number of sample coordinates depends on the complexity of the contour (edge).

$$F_o = \sum_m M_{om}, \quad m \geq 1$$

where an object is assembled by m meta objects M_{o1}, M_{o2}, …, M_{om}, $m \geq 1$.

$$F_r = \boldsymbol{\Sigma}\, O_{ij} \quad i= 1,2,\ldots,m;\ j=1,2,\ldots,k;$$

where O_{ij} is the i-th object in the j-th region of interest.

4.2 Integrating spatial relationship into SOs

On the basis of SOs, objects of interest can be represented by a set of descriptors defined above. In order to retrieve an image related to the variable locations of objects, we should integrate spatial relationship of objects into SOs. An example on spatial retrieval can be depicted by Figure 2.

In Figure 2, three objects of interest A, B, C could be

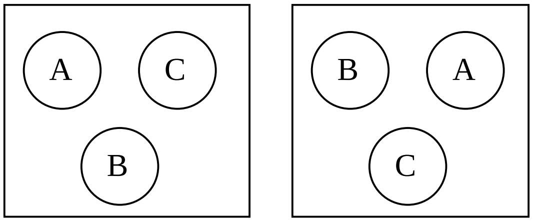

Figure 2. An example of spatial retrieval

located at different locations respectively. The locations of A, B, C can be defined by the spatial attribute of SOs. The information for locations of objects should be considered as a constraint. When we ignore the constraint, only other SOs participate in similarity matching. The changes on locations of objects do not influence the result of the similarity retrieval. This is an important specification of a visual ontology based approach.

5. The Colour Ontologies (COs)

The distribution of colours in an image can be represented in individual colour spaces. Some researchers represent the colour scheme using colour histograms. The use of colour histograms for image retrieval has been explored in [7, 8, 9]. No universal colour feature space is most suitable for CBIR using colour histogram.

One of the most important approaches for colour representation is to map a continuously distributed colour space into a set of discrete colour codes or wavelet coefficients (WCs). The approach using MWT of the images has been proposed [10]. The comparisons of significant WCs between query and target images are performed to search desired images with colour similarity. The colour space chosen for representing the images and performing the decomposition is *YIQ*. The experiments have shown dramatic improvement, in both speed and success rate, over using conventional L^1, L^2, or colour histogram norm.

Here we represent and compare the colour similarity between query and target images using both the visual COs and MWT. The spatial relationship of objects or region of interest can be represented by visual COs. We first introduce the colour spatial model for the computation of significant WCs with threshold constraints. Then we define a set of visual COs that can be used by user-computer interaction. These visual COs can also be used to represent the colour distribution of an image in an intuitive way and can be transformed into corresponding *YIQ*, *YUV*, or *RGB* colour space for wavelet based or histogram based similarity matching.

5.1. Representation of Colour feature

We use both COs and most significant WCs of the 2-D Haar MWT to represent the colour feature. A standard Haar WT of an image is simple to code. It involves a one-dimensional decomposition of each row of the image, followed by a one-dimensional decomposition for each column of the result. On contrast, non-standard Haar WT is more efficient to compute. For an $m \times m$ image, the standard decomposition requires $4(m^2-m)$ assignment operations, while the non-standard decomposition requires only 8/3 (m^2-1) assignment operations. Both standard basis and non-standard basis of the wavelet have been selected [14]. For an $m \times m$ image, there are $m \times m$ different WCs for each colour channel. Only largest magnitude coefficients should be selected in order to accelerate the search, reduce the storage, and improve the discriminatory power.

The COs we propose here is a set of discrete colour descriptors represented by common colours with rough chroma and logic relations. They can also be represented by continuously adjustable RGB colour components with rough chroma and logic relations.

5.2. Visual colour ontologies

On the basis of MWT discussed above, we define a set of visual COs for colour features. These visual COs can be integrated with the other VOs we proposed.

Table 5. Definition of the visual COs

Descriptor	Sym-bol	Descri-ption	Colour space	Chroma	Logic opera-tor
Red	C_{red}	PCC	RGB	dark, light, normal	$\wedge,\vee,\neg$
Green	C_{green}	PCC	RGB	dark, light, normal	$\wedge,\vee,\neg$
Blue	C_{blue}	PCC	RGB	dark, light, normal	$\wedge,\vee,\neg$
Rough colours	RCr	mono/ multi colour	YUV RGB YIQ	dark, light, normal	$\wedge,\vee,\neg$

In Table 5, primary colour channels are abbreviated as PCC. The rough colours include a set of common colours such as brown, yellow, green, blue, black, red, orange, pink, purple, white, etc. The VOs can also be represented in RGB primary colour channels. Every colour consists of a rough chroma attribute, dark, light, or normal, respectively.

The visual COs defined by RGB space can be represented by the continuously adjustable colour that controlled by corresponding scroll bar. Three primary colour channels can be controlled by the corresponding scroll bars respectively. Using interactive adjustment, the desired colour can be obtained. The RGB space can be further transformed to YIQ, YUV, HSV, or other colour spaces. In this way, the most significant WCs can be obtained using MWT.

5.3. Colour similarity matching

The colour similarity matching using both VOs and wavelets can be divided into two cases:

1. The colour feature represented by a set of common colours with a set of rough chroma. In this circumstance, COs should be represented by the corresponding colour descriptors. Each set of descriptors can be considered as a discrete colour distribution in a certain colour space, such as a RGB colour space. The relationship between common COs and most significant WCs can be defined in advance. The deficiency of this representation is that it is difficult to obtain any continuous distribution of colours.
2. The colour feature represented by RGB channels with a set of rough chroma. Each channel of RGB has certain value that can be easily transformed into any other colour spaces. The RGB COs can be used to represent any colour distribution. The MWT approach then can be used to obtain a set of most significant WCs.

The similarity matching is performed between query and target COs. After the submission of a real query image, which is not represented by VOs, the system first decomposes the most significant COs using MWT and is mapped into a set of query COs. Then the query COs can be matched with indexed target COs and rough chroma constraints. When a query image is interactively defined as a set of COs the colour similarity matching becomes simple.

6. The Texture Ontologies (TOs)

Textures are homogeneous patterns or spatial arrangements of pixels that regional intensity or colour alone does not sufficiently describe. Therefore textures have statistical properties, structural properties, or both [5].

Texture of images is one of the important features for image retrieval. It describes the content of many real world images: for instance, clouds, tree, bricks, hair, fabric all have textural characteristics. It is hard to adequately model texture. The investigations on texture have not yet produced any clear solutions for the problems of texture analysis, classification, segmentation and synthesis [6]. Most techniques for texture analysis are investigated and/or evaluated using Brodatz textures and are not generally applicable to "unconstrained" images. Recent attempts at modelling texture include random field modelling, fractal geometry, spatial grey-level dependencies, co-occurrence matrices and spatial-frequency techniques using Gabor filtering [9].

We describe texture of images using a set of ontologies that can be interpreted into corresponding descriptors and texture elements. A collection of texture elements is stored in the texture library. These elements should be indexed in advance and mapped into corresponding descriptors as well.

6.1. Definition of visual TOs

Here we define a set of TOs for CBIR. In order to avoid time-consuming computations in texture matching, TOs should be represented by the related texture element, descriptor, rotated angle to X axis, and rough location.

Table 6. Definition of TOs

Texture element	*Descriptor*	*Rough location*	*Rotated angle*	*Logic operator*
texture name	$T_{\text{texture name}}$	m×m areas	$0\sim 2\pi$	$\wedge,\vee,\neg$

In Table 6, texture elements consist of a set of given texture collection such as body of water, crop field, mountain, cloud, tree, hair, woven aluminum wire, reptile skin, brick wall, French canvas, plastic pellets, and so on. These texture elements come from the Brodatz texture collection. A texture element can be named using the corresponding descriptor as $T_{\text{texture name abbreviation}}$. Rough location can be represented by m×m area codes where m=4, generally. A set of logic operators, (ie. 'and', 'or', 'not'), can be added to represent the logic relations among multiple texture of subimages. Based on the consideration of individual rotated angle of the texture, the angle of the texture to X-axis should be defined in a range of $0\sim2\pi$.

6.2. Mapping TOs to WCs

We use the MWT to calculate most significant texture WCs for each texture element in the texture library. Because a texture ontology corresponds to a certain texture element, we should calculate the most significant WCs for every texture element in advance. This means that if we use the similar approach as colour WT, the TOs should be mapped into a set of most significant WCs.

6.3. Similarity matching by TOs and WCs

The similarity matching using TOs and most significant WCs can be performed in the following way:

1. Segment an overall query image into m×m subimages. Each subimage is assigned an individual area code.
2. For each query subimage perform the 2-D MWT so

that a set of most significant WCs can be obtained.

3. According to the area code which matches corresponding TOs (most significant WCs) stored in advance with rough location, rotated angle, and logic relation constraints, a similarity score can be obtained.

In the case of more complicated texture of images, more subimages should be segmented in order to obtain higher matching accuracy.

When query texture is interactively defined using the sample texture elements in the texture library the texture similarity matching becomes easy. Sometimes the TOs should be integrated with other ontologies such as shapes for more efficient retrievals.

7. Experiments

We measured the speed of our integrated VOs and 2-D MWT prototype and tested how well our approach performs as the size of the database was increased (n=550, and n=2190, respectively). We therefore gathered 2190 images from the World Wide Web with gif or jpg format. We converted these images into thumbnail and computed the corresponding WCs for each image, mapped WCs into VOs, and stored the resulting database locally. On a NT platform using extended SQL to make queries according to individual VOs. The success rate and speed of our prototype are shown as follows:

Table 7. Experiment results

Features	Success rate		Time (second)	
	n=550	n=2190	n=550	n=2190
Shape	*0.87*	*0.82*	*0.32*	*0.56*
Colour	*0.76*	*0.73*	*0.21*	*0.45*
Texture	*0.75*	*0.71*	*0.42*	*0.68*
Spatial	*0.65*	*0.62*	*0.19*	*0.28*
Spatial/texture	*0.91*	*0.87*	*0.46*	*0.75*
Spatial/shape	*0.93*	*0.89*	*0.41*	*0.65*

The experiment results in Table 7 have shown encouraging performance. We believe that further investigations on integrating VOs and MWT should lead to better results.

8. Conclusion

We have proposed an ad hoc approach for CBIR by integrating VOs and MWT. We have defined a set of meta ontologies for assembling desired VOs. The VOs are mapped into corresponding descriptors and the most significant 2-D WCs for similarity matching. Mapping from continuous feature spaces to discrete ontology spaces and a set of WCs, more efficient image contents similarity matching can be achieved in an intuitive way. The initial experiments have shown encouraging results.

Our further investigations include perfect representation of VOs, a more robust and efficient WT algorithm, the distributed multi-agent based retrieval, and an interactive visual query language.

References

[1] Stephen McCloskey and David Bell. Towards the Representation of Spatial Ontologies. University of Ulster at Jordanstown, Dec. 1997.

[2] Sabine Timpf and Andrew U. Frank. Using Hierarchical Spatial Data Structures for Hierarchical Spatial Reasoning. *Proc. of COSIT'97*, Oct. 1997.

[3] Todd Federico and Nancy Franklin. Long-Term Spatial Representations From Pictorial and Textual Input. *Proc. of COSIT'97*, Oct, 1997.

[4] Debra T. Burhans, et al. Domain Specific Understanding of Spatial Expressions. *Proc. of First Workshop on the Representation and Processing of Spatial Expressions*. August 1995.

[5] R. M. Haralick. Statistical and structural approaches to texture. *Proc. Of the IEEE*, 67(5), May 1979.

[6] M. Turceryan and A. Jain. Texture analysis. In C. H. Chen, L. F. Pau, and P. S. P. Wang, editors, *Handbook of pattern Recognition and Computer Vision*, pp 235-276. World Scientific Publishing Company, River Edge, NJ, 1993.

[7] M. J. Swain and D. H. Ballard. Colour indexing. *International Journal of Computer Vision*, Vol. 7, No.1, pp. 11-32, 1991.

[8] F. Ennesser and G. medioni. Finding waldo, or focus of attention using local colour information. *IEEE Trans. Pattern Anal. Machine Intell.*, 17(8), August 1995.

[9] J. R. Smith and S. –F. Chang. Tools and techniques for colour image retrieval. In *Symposium on Electronic Imaging: Science and Technology – Storage & Retrieval for Image and Video Database IV*, volume 2670, pp. 426-437, San Jose, CA, February 1996. IS&T/SPIE.

[10] Charles E. Jacobs, Adam Finkelstein, and David H. Salesin. Fast Multiresolution Image Querying. *Proceedings of ACM SIGGRAPH*, New York, 1995.

[11] A. Grossman and J. Morlet. Decompositions of hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.*, 15:723-736, 1984.

[12] I. Daubechies. Orthonormal bases of compactly supported wavelets. Comm. *Pure Appl. Math.*, XLI:909-996, 1988.

[13] John R. Smith. Integrated Spatial and Feature Image Systems: Retrieval, Analysis and Compression. Ph. D thesis, Columbia University, 1997.

[14] G.Beylkin, R.Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44(2):141-183, March 1991.

[15] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Patt. Anal. Mach. Intell.*, 11(7):674-793, 1989.

[16] V. Ogle and M. Stonebraker: Chabot: Retrieval from a relational database of images, *IEEE Computer*, vol. 28, pp. 40-48, September 1995.

[17] Myron Flickner, et al.: Query by image and video content: the QBIC system, *IEEE Computer*, 28(9):23-32, September 1995.

[18] A. Pentland, et al.: Photobook: Content-based Manipulation of Image Databases, *M.I.T. Media Laboratory Technical Report*, June 1995.