In-network content based image recommendation system for Content-aware Networks

Marta Barrilero, Silvia Uribe, Maria Alduan, Faustino Sánchez, Federico Alvarez
Visual Telecomunications Applications Group (G@TV)
Universidad Politécnica de Madrid, UPM
Madrid, Spain
{mbg, sum, mam, fsg, fag}@ gatv.ssr.upm.es

***Abstract*— This paper describes a novel content-based image recommendation system based on new image low level descriptors derived from the well known MPEG-7 parameters. Furthermore, it also proposes the integration of this recommendation system into a content-aware network architecture to enhance and enrich the content delivery and improve user's experience.**

Keywords- content-based recommendation, MPEG-7 descriptors, low level image characteristics extraction, content-oriented network, content-aware network.

I. INTRODUCTION

In recent years, the amount of multimedia content on the Internet is increasing considerably, becoming very hard for most users to cope with such a great deal of content. In fact, one of the reasons of the amount of multimedia increasing content existing on the Internet is the increasing capacity of the recent networks to managing, storing and distributing more and more content, and that is why it is possible to establish the denomination "Internet of Contents", which represents the change of direction from the location issues, typical of classic networks, to the remarkable importance of the content as a native element, typical of the new Content Oriented Networks (CON). The need of content discrimination and selection is becoming a crucial task on the multimedia content managing. That is why content aware networks are becoming more and more personalized according to users characteristics and preferences, due to their design and functions are implemented depending on individual users' preferences or behaviors.

Due to the great deal of visual content on the Internet, image retrieval systems are the aim of a considerable deal of researches. In this paper a Content Based Image Recommendation System is proposed. This system is implemented over a Content-Aware Network (CAN), which allow users to search and access the content, reinforcing the network user centric approach. This CAN utilizes "content nodes" which manages the content and its metadata, which are stored into a so called Multimedia Component, that interacts with the network and the content so that the content metadata can be managed and the description of the media objects can be enriched, which is useful for the recommendation task.

The image recommendation system proposed on this paper is content based. As the main idea of content based recommendation is related to the affinity of contents preferred by users to the possible recommended contents, low level features of images have been worked with in this paper. MPEG-7 parameters related to luminance, texture and chrominance have been utilized to represent every image by means of its parameters values. Images have been identified with a vector of values, in order to be able to compare them considering other factors related to user preferences.

The reminder of this paper will be organized as follows: Section II briefly reviews predominant approaches to recommendation systems, and compares the content-based recommendation previous solutions to our system proposal in order to point out its advantages. Section III describes our new content based recommendation designed process, divided into the main steps, and their principal functions. Section IV presents the novel parameters proposed for the recommendation algorithm. Recommendation subsystem integration into content aware networks is explained in Section V and finally, Section VI covers the conclusions derived from our study.

II. RELATED WORK

As it is explained in [1], recommendation systems have became an important research area since the first researches about collaborative filtering in the earliest '90. In fact nowadays these systems are still in progress in order to find more efficient methods to study and model users' behavior and new ways to join additional data (like contextual information, user consumption data…) that make the recommendation process more effective.

Content recommendation is usually reduced to the estimation of a rating of the presented items in order to offer the most suitable content for each user. To that end, according to [2], three basic recommendation methods are available:

1) Content-based recommendation method

This method, as it's explained in [1] and [3], is based on the similarity between items characteristics and the user profile information, in order to select similar content according to the user's preferences. The user's profile contains both implicit and explicit information and it is updated in a dynamic way thanks to an implicit learning method, usually based on supervised techniques like genetic algorithms [2].

The main advantage of these systems is the possibility of new items recommendation although there are no ratings available from other users. In the same way, the main disadvantages are on one hand that it is necessary a complete user profile and a complex description of the items for a right execution, and on the other an over specialization, that is, content-based

978-1-4244-9920-5/11/$26.00 ©2011 IEEE

recommendation systems tend to recommend similar items to the consumed ones.

2) Collaborative recommendation method

In this case, the content recommendation is based on the items rating made for other users previously, according to heuristic or probabilistic methods [3] and [4]. In fact, there are two aspects taking part in the recommendation process:

- Users' affinity: considering only the ratings made by similar users.
- Items' affinity, according to users' ratings.

The main disadvantage of this method is the so called *cold start* problem [5], preventing the new items recommendation, due to the lack of previous information.

3) Hybrid recommendation method

This method takes advantage of the previous ones by means of their combination, which can be implemented in two different ways, as it is explained in [1]:

- By combining the content-based and the collaborative predictions obtained separately.
- By adding some content-based characteristics to a collaborative recommendation system.

Both solutions try to avoid the disadvantages of previous methods by improving their positive aspects.

The content based recommendation system proposed in this paper represents a new evolution of these solutions, due to the incorporation of new content descriptors in charge of characterizing, in a more effective way, the content able to be suggested, avoiding the specialization problem, and allowing its later integration in a hybrid recommendation system. This can be performed weighting the methods according to the available information (users' information in case of collaborative methods or item's information in case of content-based methods). Besides, this system is conceived to be included in the new Content-Aware Networks, taking advantage of the synergies of this new kind of network architecture.

III. Content Based Recommendation Process

The proposed image recommendation system is content-based, and it is exclusively based on low level image properties corresponding to MPEG-7 parameters. Three kinds of low level characteristics have been extracted: luminance, texture and chrominance. Software implemented by [6] has been used. Descriptors utilized are specified on next section.

There are two kinds of images: unknown images and known images. Known images have been viewed and rated by the user, so there is information about them. Only rated positively known images are going to be utilized. Unknown images are those which are going to be recommended to the user.

The proposed recommendations process (Fig. 1) is carried out by means of three steps: user preferences analysis (blue figures), unknown images processing (green figures) and final recommendation (red figure).

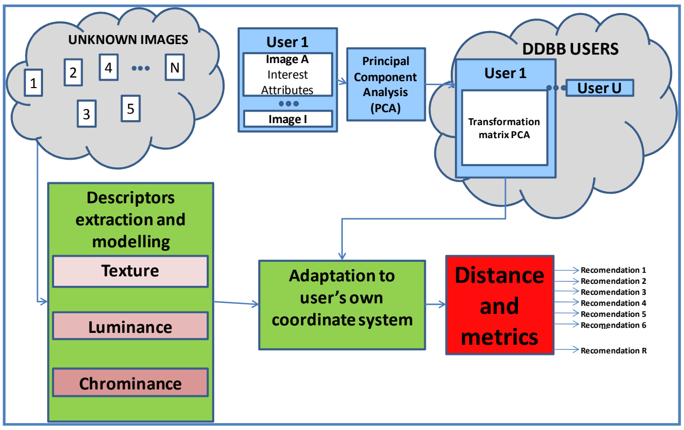

Figure 1. Content-based recommendation process

1. **User preferences analysis**: a user preferences analysis which consists of the extraction of the more relevant parameters to each user is executed.

   1.1. *Low level characteristics extraction*: using a software application, known images parameters are extracted.

   1.2. *Attribute extraction*: each image is represented by one vector whose components are the image's attributes values, related to low level characteristics (texture, luminance or chrominance). A matrix M of dimension Nxa is built with this information. N is the number of known images and a the number of evaluated attributes on each image. In this case $a = 14$. The meaning of each value will be specified on Section IV.

$$v = (\mathrm{E}_1, H_1, L_c, L_y, \mathrm{e}_1, e_2, e_3, e_4, \mathrm{N}, \mathrm{VC}_{\mathrm{intra}}, \mathrm{VC}_{\mathrm{inter}}, \mathrm{SC}, S_m, S_y)$$

   1.3. *Principal Component Analysis (PCA)*: in order to reduce the dimension of matrix M, as well as selecting the relevant factors for the analyzed user a principal component analysis is executed. PCA algorithm consists of next steps:

   a) Mean-zero transformation: mean value from each column on M is subtracted, so the columns on new matrix M' have mean-zero.

   b) Covariance calculation: the covariance of M' is calculated and a matrix of dimension axa is obtained: V. This matrix represents the variance between columns on M'.

   c) Eigenvectors calculation: eigenvectors of V are calculated and the matrix T (dimension axa) is obtained. T will be the *transformation matrix* and it will be necessary multiplying by this matrix to change the coordinate space.

$$Coordinates = M' * T$$

   The dimension of coordinates matrix will be Nxa but its columns will be sorted from more to less variance. This means that the first columns will be the most representative and will be enough to represent the vectors without much loss of precision. For example, the first three columns of the matrix allow to represents N images with three dimensions. So, after executing PCA the matrix input can be transformed into a less-dimension matrix which brings similar parameters together on the same dimension. This is very useful because the parameters which have similar importance for the user are brought together, so the matrix will model user preferences.

1.4. *Transformation matrix storage*: matrix T is stored into a database and subsequently applied to the unknown images vectors in order to change them into a personalized dimensional space for every user. This matrix will represent user preferences.

2. **Unknown images processing**: the main idea of this step consists of extracting unknown images properties and applying them the same linear transformation (matrix T) applied to known images previously, in order to being able to comparing them and recommending the most-similar to preferable images.

   2.1. *Low level characteristics extraction*: the same process as in step 1.1 is carried out, but with unknown images.

   2.2. *User matrix execution:* matrix T is applied now to each of the unknown images in order to changing the coordinate space.

3. **Recommendation**

   3.1. *Distance calculation:* since both vectors which represent known images and unknown images have gone through the same changes, both have been transformed into the same space, so the distance between both known images and unknown images can be calculated.

   3.2. *Best images selection:* according to Mahalanobis distance, the closest images are selected and recommended to the user.

IV. IMAGE RECOMMENDATION DESCRIPTORS

As it was explained before, the recommendation system shown in this paper makes use of the low level metadata of the content to perform the suggestion process, without taking into consideration other kind of information such as semantic annotation. For this reason, the selected low level parameters have to provide enough information to allow an optimum recommendation process that fulfill the users' expectations and preferences.

Regarding this, the MPEG-7 standard defines and provides a set of descriptors for visual media [7] which are usually used in content based retrieval systems. These descriptors are divided into five groups (color, texture, shape, motion and others), and each of them consists of a feature extraction mechanism, a description (in XML and binary format) and a set of guidelines that indicate how to apply each descriptor on different kinds of media.

Since our system works with images, only color and texture MPEG-7 descriptors have been considered, avoiding the motion (we are not recommending picture in motion) and the shape ones (we are only using low level descriptors). They are obtained by the MPEG-7 Low Level Feature Extraction command line tool [8]. Besides, other descriptors have been derived from the previous ones in order to facilitate and optimize the recommendation process. This new parameters can also be divided into three different kinds, according to the analyzed characteristic.

A. Texture descriptors

In this case, the designed parameters derive from the MPEG-7 Edge Histogram Descriptor (EHD) [7], which is in charge of establishing the edge distribution along the image by the obtaining of local edges with different orientations in subimages.

1) Line energy

This parameter (E_l) is in charge of measuring the total line density in the image according to its line energy level, in order to distinguish between images with no transitions and full edges images.

$$E_{lTot.}^2 = E_{45^\circ}^2 + E_{90^\circ}^2 + E_{135^\circ}^2 + E_{180^\circ}^2 + E_{Other}^2 \qquad (1)$$

where E_i shows the image edge distribution along each direction.

2) Line homogeneity

This descriptor (H_l) is in charge of establishing the image line continuity. It divides the image into 16 subimages and then calculates the distribution line variance between 4x4 subimages neighbor blocks as follows:

$$H_1 = V_1 + V_2 + V_3 + V_4 + V_5 \qquad (2)$$

where

V_1= variance between (1,1),(1,2),(2,1),(2,2) subimages

V_2= variance between (1,3),(1,4),(2,3),(2,4) subimages

V_3= variance between (3,1),(3,2),(4,1),(4,2) subimages

V_4= variance between (3,3),(3,4),(4,3),(4,4) subimages

V_5= variance of V_1, V_2, V_3, V_4.

and taking into account the line distribution along each direction.

3) Entropy variety

This parameter represents the variance of entropy along the whole image, and it is calculated dividing the image into a set of 5x5 subimages and obtaining the entropy of each one of them. The final parameter is the variance of the 25 subimages

B. General Luminance descriptors

These parameters are in charge of determining the saturation of the content, according to different aspects of the image.

1) Direct luminance

This parameter derives directly from the MPEG-7 Color Layout Descriptor (CLD) [7], which represents the color spatial distribution of the image obtained by the application of the Discrete Cosine Transform (DCT) to the image colors in the YCrCb color space.

Direct luminance is in charge of obtaining the saturation mean level of each image, in order to distinguish between clear and dark content. The descriptor is composed of two values (L_c,L_v): on one hand the mean of the luminance and on the other its variance, which represents luminance dispersion.

2) Bit-plane distribution entropy

This descriptor establishes the bit-plane distribution entropy according to [9]: taking into account a gray-stone image, each element is an integer lying between [0,255]. This value establishes the intensity value of each pixel, which can be represented by an 8 bit binary vector, one for each plane ($b_7,b_6,b_5,b_4,b_3,b_2,b_1,b_0$). In order to avoid small variation of intensity affecting all bit planes, each pixel is expressed by a Gray-code (which provides more significant information), adopted as follows:

$$g_i = \begin{cases} b_i \oplus b_{i+1}, & 0 \le i \le m-2 \\ b_i, & i = m-1 \end{cases} \quad (3)$$

where $\oplus$ denotes XOR operation, b_i is the *ith* bit-plane and g_i is the *ith* bit plane expressed by Gray code. After that, we consider the entropy of the four highest bit-planes, which contain most of the structural information in the image.

Therefore, this descriptor is composed by four components, one for each bit-plane entropy (E_1, E_2, E_3, E_4).

C. Color descriptors

1) Chromatic variety

This parameter derives from the MPEG-7 Dominant Color Descriptor (DCD) [7], and is in charge of analyzing the color variety along the image, according to the detected dominant colors.

Chromatic variety descriptor is composed by three values (N, VC_{intra}, VC_{inter}): one indicating the number of dominant colors (extracted directly from DCD), other establishing the variance between different values of the same color, and finally the variance between each color (intra-chromatic and inter-chromatic variety).

2) Spatial coherence

This descriptor also derives directly from MPEG-7 DCD descriptor and its main objective is to determine the continuity of the color in the image. It is composed by only one component.

3) HSV color space

By making a color space change, we select the color values and its saturation on the image, in order to establish their impact in the user selection. This parameter is composed by two values (S_m, S_v): on one hand, the saturation mean value and on the other the saturation variance.

4) Color planarity

This descriptor represents the homogeneity on each dominant color in the image in terms of color hue. For example, an image with the same two colors will have much more color planarity than another one with many different color hues.

It is calculated as follows: the image is divided into pixels blocks so that each block is a square with a pixel-side equal to the minimum side of the image divided into a selected factor (in the tests the factor is 8). The most frequent color in each block is selected and the percentage of pixels with a similar color (with a difference of 20 in the RGB system) exclusively in that block is calculated. The general color planarity will be the mean of all percentages.

V. CONTENT BASED IMAGE RECOMMENDATION SUBSYSTEM IN CONTENT-AWARE NETWORKS

Internet evolution includes new paradigms to offer better content delivery services to the users, who give more value to the content meanwhile the Internet, due to its original conception, gives more value to the localization of that content. For this reason, a major change is needed in the "Internet of Contents" orientation: from "where" to "what". This new paradigm is called Content Oriented Networks (CON), which addresses the basic needs of the Internet design to cope with the content as a native element.

The Recommendation System described in this manuscript can be implemented at network level over a Content-Aware Network (CAN) avoiding ad-hoc solutions.

One of the main advantages of the CANs is that they provide the media content with searchable and accessible capabilities. The Recommendation System, by taking advantage of these capabilities, reinforces the network user centric approach.

To implement the Recommendation System over a CAN, it can be allocated within the network cloud. In this scenario, the content is managed by "Content Nodes" being its associated metadata stored into a so called Multimedia Component, that allows the network to access and route both media essence and metadata information. This Multimedia Component provides an enriched description of the media objects since it is continuously enriched by the recommender.

The Multimedia Component stores different both low-level and high-level metadata types, some of them used by the Recommendation System not only to carry out the recommendation but also to generate new affinity metadata. These Affinity Metadata model the subjective media content perception of a user; their main objective is to establish affinity relations between users and multimedia objects. For instance, they include the measurement of the inherent geometry of a picture or some composite rules, which can be influenced by the perception of a user.

Fig. 2 depicts a high level approximation of the Recommendation System implemented over a CAN. This proposed approximation is "protocol agnostic" (messages, naming…) so it could be adapted to any of the Content Centric Networking (CCN) protocols already proposed [10].

The multimedia search process shown in Fig. 2 consists of 4 steps:

(1) First, the process is initiated with a search query containing its corresponding metadata.

(2) The metadata are flooded to the CAN reaching every “Content Node”. When a “Content Node” receives a metadata search query, a search over the *Multimedia Component* (MC) takes place in order to find the images that satisfy this request. After this operation, every “Content Node” returns, as a reply, the metadata associated to the images.

(3) The reply is sent to the *Recommendation System* by the “CAN routers” to obtain a list of the recommendations over this result. To perform this, the Recommendation System uses both the received image metadata and the user information that is stored into the user profiles (located within the Recommendation System.)

The *Multimedia Component* allows the existence of a highly sophisticated content-based recommender module, since a big amount of information is available for every multimedia element. The CAN contributes to the knowledge of the multimedia objects users’ consumption, due to the fact that these objects can be unambiguously identified within the network. This knowledge allows the development of automatic and transparent content-based recommendation algorithms.

(4) Finally, on one hand the new metadata generated by the Recommendation System sent back to the Multimedia Component in order to enrich the description of the corresponding content and in the other hand, the list of recommended images are retrieved to satisfy the initial query.

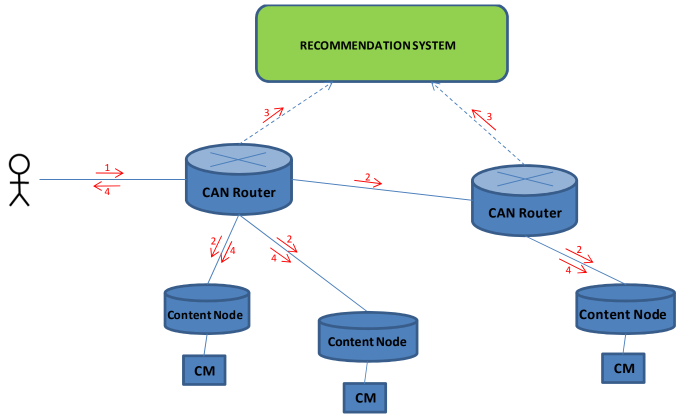

Figure 2. Recommendation system in the multimedia search process

VI. EXPERIMENTAL RESULTS

To evaluate the proposed system 200 images were selected from the virtual art gallery *Ciudad de la pintura* [11]. These images represent paintings from all centuries and were scored by 60 users with a scoring from 1 (dislike) to 5 (like).

For each user, the scoring for 150 images was selected. Then, images from this dataset with a major score were considered (4 or 5) as preferred known images, while the other 50 images could represent unknown images. It is important consider that not all users have the same general tastes, since there can be users with a considerable preference for most of paintings, along with users with none preference for hardly any painting. To deal with this, the mean of all scores from each user has been taken into account.

Several tests have been implemented to evaluate the system. In this paper two of them are explained.

1) The first test is useful to check that the defined “Image recommendation descriptors” characterizes in a good way the tastes of the users. It also establishes a classification from best to worst characterization descriptors.

Let *a be a generic descriptor*. The measurement of the *Importance* of *a* has been developed comparing the variance of the values of the descriptor in the 50 unknown images to the variance of the values of the descriptor in the recommended images. The *Importance* of descriptor *a* can be defined as follows:

$$Importance_a = \sum_{user=1}^{60} \frac{Var(X_a)}{Var(Y_a)} \tag{4}$$

where the values of the descriptor in the 50 unknown images belong to X, and the values of the descriptor in the recommended images belong to Y.

Other defined parameter is *Affected Users*, which establishes how many people from the 60 tested users are influenced by each descriptor. The way to obtain the final value is to add how many users fulfill the condition $Var(X_a)/Var(Y_a) > threshold$. The tests of Table I have used the value threshold=1.15. The descriptors are sort from best to worst.

TABLE I. IMAGE RECOMMENDATION DESCRIPTORS EVALUATION

Descriptor type	***Descriptor***	***Importance***	***Affected users***
Luminance descriptors	Entropy variety	213.79	22 (37%)
Texture descriptors	Line homogeneity	160.2951	44 (73%)
Color descriptors	Chromatic variety (intra)	138.0416	43 (72%)
Luminance descriptors	Bit-plane distribution entropy (4)	100.0674	56 (93%)
Texture descriptors	Line energy	91.4375	45 (75%)
Color descriptors	Color planarity	74.1061	40 (67%)
Color descriptors	HSV color space (saturation variance)	73.4121	36 (60%)
Luminance descriptors	Bit-plane distribution entropy (1)	68.5153	41 (68%)
Color descriptors	HSV color space (saturation mean)	63.7606	44 (73%)
Luminance descriptors	Direct luminance (variance)	60.5754	37 (62%)
Color descriptors	Chromatic variety (inter)	44.1947	31 (52%)

Descriptor type	*Descriptor*	*Importance*	*Affected users*
Color descriptors	Spatial coherence	41.7862	26 (43%)
Luminance descriptors	Bit-plane distribution entropy (3)	27.8353	21 (35%)
Color descriptors	Chromatic variety (N)	20.6617	16 (27%)
Luminance descriptors	Bit-plane distribution entropy (2)	17.2227	12 (20%)
Luminance descriptors	Direct luminance (mean)	5.3371	4 (7%)

2) For the second test, an Improvement measurement has been defined. The recommendation system has selected the n first recommendations and its mean was calculated. Then it has been compared to the mean of all scores from one user. The improvement of the recommendation is considered as the increase of the recommendation mean over the user mean.

The results have been obtained modifying the number of recommended images from 1 to 50 recommendations, which is the total number of unknown images. Therefore, if the system recommends 50 images, the improvement should be 1, because there is no recommendation, and this value should grow up while the system is recommending fewer images, which should be the best images for the user.

Figure 3 shows the results of the mean Improvement value for the 60 users corresponding to different number of recommended images.

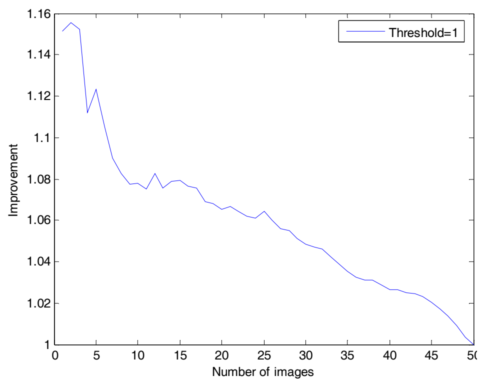

Figure 3. Mean Improvement value

VII. Conclusions

A new system to provide personal image recommendation according to user's implicit information is described in this paper. This solution presents three key innovations beyond current similar systems:

1. It develops new low level descriptors, based on the combination of MPEG-7 parameters. These new descriptors represent a more efficient way of providing representative information about users' content preferences in an implicit manner.
2. It describes a new image content-based recommendation process, whose main objective is to provide users with the right content according to their preferences and consumption behavior in a more effective way, since it does not depend on semantic tags, as in current content-based recommendation systems, but on formal and aesthetic characteristics, independent of the image topic.
3. It enriches the content delivery and reinforces the network user centric approach by the integration of this recommendation system into the content aware network as a new kind of functionality.

Acknowledgment

This publication is based on work performed in the framework of the Spanish national project BUSCAMEDIA (CEN-20091026), which is partially funded by the CDTI - Ministry for Science and Innovation. The authors would like to acknowledge the contributions of colleagues from the partners of the BUSCAMEDIA project.

References

[1] G. Adomavicius, A. Tuzhilin. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions". IEEE Transactions on Knowledge and Data Engineering, Vol. 17, No. 6. June 2006. Pages 734-742.

[2] B. Sheth, P. Maes. "Evolving Agents for Personalized Information Filtering". Proceedings to the Ninth Conference on Artificial Intelligence for Applications. March 1993. Pages 345-352.

[3] M. Balabanovic, Y. Shoham. "Fab: Content-based, collaborative recommendation". Communications of ACM 40,March 1997, 66-72.

[4] L. Candillier , F. Meyer , M. Boullé. "Comparing State-of-the-Art Collaborative Filtering Systems". Proceedings of the 5th international conference on Machine Learning and Data Mining in Pattern Recognition, July 18-20, 2007, Leipzig, Germany.

[5] A. Barragáns, J. J. Pazos, A. Fernández, J. García, M. López. "What's on tv tonight? An efficient and effective Personalized Recommender System of TV Programs". IEEE Transactions on Cosnumer Electronics. Vol. 55, No. 1, Feb. 2009. Pages 286-294

[6] M. Bastan. October 2009 Bilkent University Department of Computer Engineering Bilkent, Ankara, Turkiye

[7] B. S. Manjunath, P. Salembier and T.Sikora "Introduction to MPEG-7: Multimedia Content Description interface". Ed. John Wiley & Sons, Ltd. 2002.

[8] M. Bastan, H. Cam, Ugur Gudukbay, Ozgur Ulusoy, "An MPEG-7 Compatible Video Retrieval System with Integrated Support for Complex Multimodal Queries,"IEEE MultiMedia, 2009.

[9] Zhao Shan, Wang Hai-tao, "Image Retrieval Based on Bit-plane Distribution Entropy" 2008 International Conference on Computer Science and Software Engineering.

[10] V. Jacobson, D. K. Smetters J. D. Thornton, M. F. Plass, N. Briggs, R. Braynard, "Networking named content", Proceedings of the 5th ACM International Conference on Emerging Networking Experiments and Technologies (CoNEXT 2009), 2009 December 1-4; Rome, Italy. NY: ACM 2009, pp. 1-12.

[11] http://pintura.aut.org