韓國 컴퓨터情報學會 論文誌
第13卷 第2號, 2008. 3.

2008-13-2-2-2

인간의 공감각에 기반을 둔 색청변환을 이용한 영상 인식

신 성 윤*, 문 형 윤**, 표 성 배***

Image Recognition Using Colored-hear Transformation Based On Human Synesthesia

Shin Seong Yoon*, Moon Hyung Yoon**, Pyo Seong Bae***

요 약

본 논문에서는 공유 비전과 특수한 청각에 의해 감지되는 인간의 공감각 특징을 구별하는 색청 인식을 제안한다. 카메라를 통한 시각적인 분석이 인간의 구조화된 사물 인식에 영향을 주는 것이 가능하다는 점이다. 그래서 시각장애인들이 실제 사물과 유사한 비전을 느낄 수 있도록 하는 방법에 대해 연구해왔다. 우선 특정 장면을 대표하는 영상 데이터에서 객체의 경계가 추출된다. 다음으로, 이미지에서 객체의 위치, 색상 평균 감성, 각 객체의 거리 정보, 그리고 객체 영역의 범위와 같은 4가지 특징을 추출하고, 이들 특징들을 청각적 요소로 사상한다. 청각적 요소는 시각장애인을 위한 시각 인식 형태로 제공된다. 제안된 색청 변환 시스템은 보다 빠르고 세부적인 인지 정보를 제공하고 동시에 감각을 위한 정보를 제공한다. 따라서 이 개념을 시각장애인의 영상 인식에 적용할 경우 보다 좋은 결과를 얻을 수 있다.

Abstract

In this paper, we propose colored-hear recognition that distinguishing feature of synesthesia for human sensing by shared vision and specific sense of hearing. We perceived what potential influence of human's structured object recognition by visual analysis through the camera, So we've studied how to make blind persons can feel similar vision of real object. First of all, object boundaries are detected in the image data representing a specific scene. Then, four specific features such as object location in the image focus, feeling of average color, distance information of each object, and object area are extracted from picture. Finally, mapping these features to the audition factors. The audition factors are used to recognize vision for blind persons. Proposed colored-hear transformation for recognition can get fast and detail perception, and can be transmit information for sense at the same time. Thus, we were get a good result when applied this concepts to blind person's case of image recognition.

▸ Keyword : Colored-hear Recognition, Synesthesia, audition factor

---

• 제1저자 : 신성윤
• 접수일 : 2008. 2. 18, 심사일 : 2008. 2. 20, 심사완료일 : 2008. 3. 1.
* 군산대학교 컴퓨터정보공학과 교수, ** LG CNS 차장, *** 인덕대학 컴퓨터소프트웨어과 교수

I. 서 론

In this paper, we propose conversion of between one's vision data and hearing data based on psychological cross sense on the human computer interaction. It called synesthesia conversion that shared sensation by common brain working theory. In this kind of phenomenon was discussed for a long time. This studies range over many subjects.

The concept of this paper is fast recognition by synesthesia for blind persons. Vision recognition system base on real time is not efficiently for HCI. Because we have to give a definition of each objects by form of words. That is too risky for understanding or not very efficient to make imagination about real world. Thus, Recognition message using vision need to be redefine to which another sensation stimulus not the words. Synesthesia can be one of solution in this problem. So we suggest conversion system between color and hearing using synesthesia in terms of changing notice message for vision recognition.

First, We describe concept of synesthesia. And next section, organize Colored-hear Conversion System(CCS) based on human synesthesia theory. Fourth section, we've experimented image recognition for blind person using CCS. Finally, we give you experimental result and our conclusion

II. Synesthesia

Synesthesia is a neurological condition in which two or three, or more bodily senses are coupled.[5][6][9] In one common form of synesthesia, known as music-color synesthesia, music or sounds are perceived as inherently colored.[4]

Synesthesia can occur between nearly any two senses or perceptual modes. Given the large number of forms of synesthesia, researchers have adopted a convention of indicating the type of synesthesia by using the following notation x → y, where x is the "inducer" or trigger experience, and y is the "concurrent"or additional experience. For example, perceiving letters and numbers(collectively called grapheme) as colored would be indicated as grapheme → color synesthesia.[7] ⟨Fig. 1⟩ demonstrates the logic of one test used to demonstrate the reality of synesthesia.[8][2]

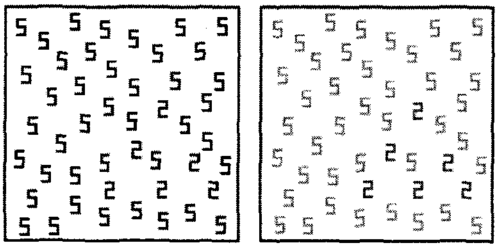

그림 1. 공감각 경험의 현실을 논증하기위해 사용되는 실험의 예
Fig. 1 An Example of A Test used to demonstrate The Reality of Synesthetic Experience

Similarly, when synesthetes see colors and movement as a result of hearing musical tones[3], it would be indicated as tone → (color, movement) synesthesia. In these cases shown that some different senses are analyze into same mechanism. In music → color synesthesia, individuals experience colors in response to tones or other aspects of musical stimuli(e.g., timbre or key). And consistent trends can be found, such that higher pitched notes are experienced as being more brightly colored.[1]

The presence of similar patterns of pitch-brightness matching in non-synesthetic subjects suggests that this form of synesthesia shares mechanisms with non-synesthetes.

In our system has color → sounds model by shared mechanism on the basis of synesthesia as music-color synesthesia. Elements of sounds is correspond converse music and color elements corresponding to other hands.

III. Colored-hear Conversion System

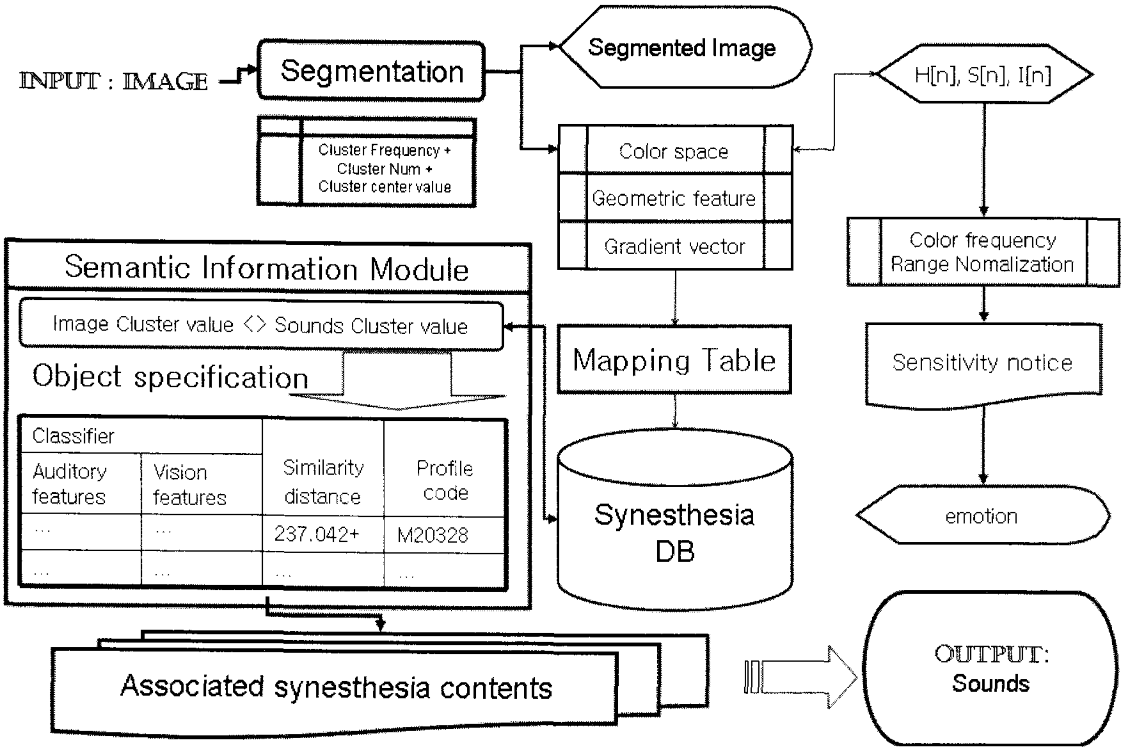

그림 2. CCS 모델
Fig. 2 CCS Model

3.1 Synesthesia space modeling

The synesthesia conversion system should create sense space using specific object features. Vision recognition has quite complex and important features. It needed to process of image feature extract from each object such as color, pattern, contour for semantic information. But we can extract object simply using distance of colors. The color is typical of object segmentation. Each segmented object can signify to features such as where or what it is. The synesthesia's sense space is consists of these each objects feature parameters.

〈Fig. 2〉 gives you an overview of this system. The Color is represents perception vision features and first part of Diagram concerned separate objects. Once the object models are segmentalized, those are then used to normalize the images and extract parameters that specific object features and It can be used to convert the auditive features. First, Image features builds into a Image cluster value such as specific color, geometric coordinates and gradient vector. Every Image cluster values are compare with Sound cluster values in the Synesthesia Database. Sounds cluster values are already defined by synesthesia phenomenon. Then, Every single Image cluster values are convert to sound cluster values with similarity distance and We can get another object specification space. Eventually, the image space objects are recognized using auditive features as user defined sounds.

3.1.1. Image segmentation via clustering

There are lots of classifier methods that could be used to purpose of image segmentation. The *K*-means algorithm is an algorithm to cluster objects based on attributes into *k* partitions. And it is typical method for color image segmentation. It is similar to the expectation-maximization algorithm for mixtures of Gaussians in that they both attempt to find the centers of natural clusters in the data. It assumes that the object attributes form a vector space. The objective it tries to achieve is to *minimize total intra cluster variance, or, the* squared error functionwhere there are k cluster S_i , i =1, 2,…, k and μ_i is the centroid or mean point of all the points $x_j \in S_i$.

$$V=\sum_{i=1}^{k}\sum_{x_j\in S_i}|x_j-\mu_i|^2 \qquad (1)$$

In addition, It can create auto-adaptive initialization for number of K and average values using distributed color space. Determined number of K and average values are calculate from previous methods. It then calculates the mean point of each image points set. Every mean point of objects calculated distances and group cluster contain near point. And then renewal average value that mean point of repeat process. Clustering and average process will run until it's not change the average values. Eventually, we can get the object area and sequence of objects units.

3.1.2. Feature Extraction

Once the objects are segmented, each object regions are cropped out and analyzed to recover the object's feature factor. For the purpose of conversion of color, we need to extract the objects feature factors robustly and efficiently. Also, rather than just extraction the color of object features, we need to positions of objects and shape size of object. Feeling of average color is corresponding to centric point of cluster space. And HIS color Model can transmit feeling to human more easily.

Thus, color model of centric point of object clusters are change to HIS color space. Object location in the image corresponding to geometric feature as coordinate.

Assume, that we have a focus blocks set of segmented object region $\{F_1, F_2, \ldots, F_n\}$ where each focus block F_k is pre-annotated with a corresponding vector of shape parameters S_k. The vector of shape parameters S is a stack of x, y coordinates of fiducial points. And Object area size E_k *corresponding to each cluster's space size.*

$$f(x)=\sum_{i=1}^{m}\sum_{k=1}^{n}F_k(S_k,E_k)+V \qquad (2)$$

Distance information of individual focus block are affects contents label. Contents label is number of clusters.

3.2 Definition of auditive factors

A study in the Ward, J.[1] says sounds can convert to colors using synesthesia. In other words, there are relations when we recognize from which senses as vision or sounds. And it can be change conversely like from colors to sounds. So segmented objects can corresponding to auditory definition.

⟨Fig. 3⟩ shows description of vision factors and audition factors for Colored-hear system. Audition factors are consist of such as Octave, Pitch, Velocity, Gain, Position, Timbre. Octave means range of sounds. And pitch is sounds level at a higher or lower within a possible range.

Velocity can make feeling speed of sounds. Gain is same called sound volume. And the last of auditory factor, Position is the source of sounds coordinate

3.3 Feature Space Mapping-Table

Once the factor parameters are extracted from image the next step is to mapping the auditory factors they corresponding to. Our system's feature mapping table has linear combination between each connected factors For example, assume that hue range of any cluster centric points is under 30. It is corresponding to pitch range 1.7, also it's close to area of "mi". The details of mapping table are shown

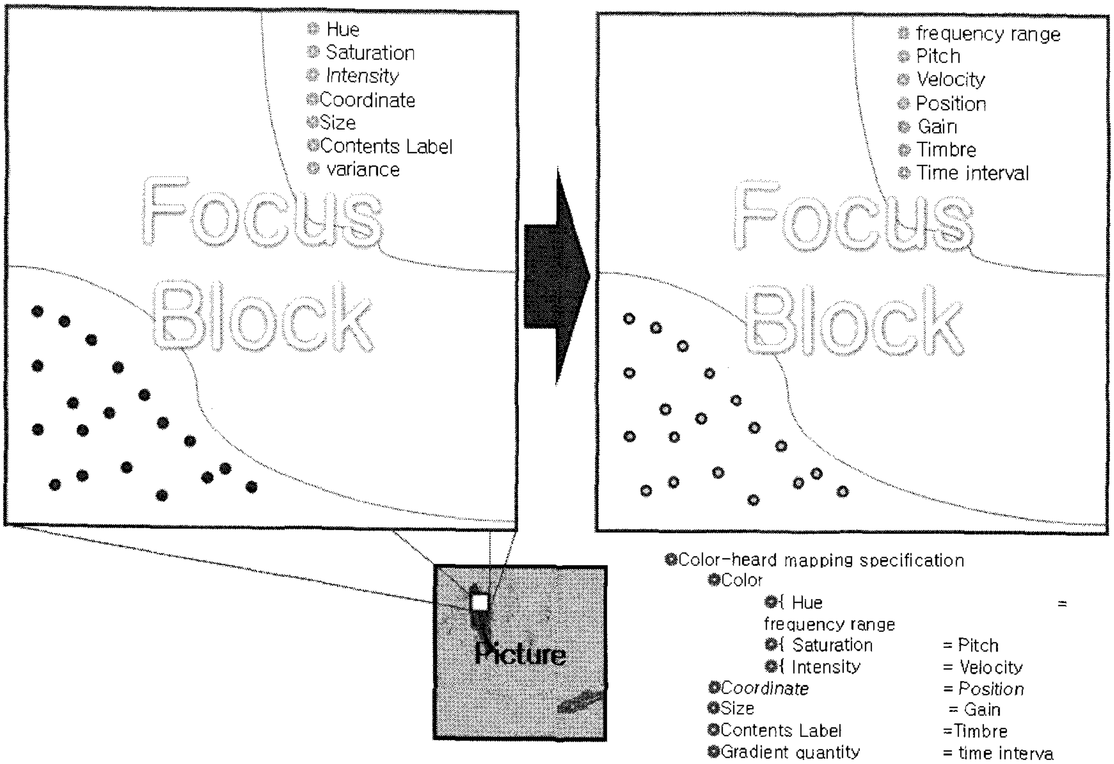

그림 3. 색청 요소 기술의 공감각
Fig. 3 Synesthesia of Colored-hear Factors Description

in ⟨Table 1⟩. Every single centric points of color clusters are connected to certain auditive elements.

표 1. 특징 공간 매핑 테이블의 예
Table 1. Example of Feature Space Mapping Table

Hue	Saturation	Intensity	Size
30	120	130	6.6
Coordinate	Contents Label	Pitch	Octave
X, Y, (Z)	20	1.7(mi)	3.0
Velocity	Gain	Position	Timbre
30	8	X, Y, (Z)	E. Piano 3

In brief, the synesthesia conversion is first processed to separate the object areas. During the clustering the focus block of object area are projected on the corresponding auditive factor space. These projections are used to mapping the sounds factors as explained above.

IV. Experimental Result

The system was evaluated on two data sets. The first dataset is completely natural that unannounced about synesthesia and relation of Colored-hear. The results are indicative of how a system like this would work in real-world applications. The second database we use CCS database given by synesthesia theory. Also we give the information that Colored-hear relation to blind persons and auditive features are using each 20 pieces sounds for different timbre. The CCS recognition database has 3 blind persons in a real situation. A vision camera was placed on top of the monitor and multi-channel sound system was placed on the around target. ⟨Table 2⟩ shows result for synesthesia combinations factors given by 40 pictures.

표 2. 색청 결합 요소들을 위한 결과
Table 2. Results for Synesthesia Combinations Factors

Combination Factors	Correct Combination	Misses	Average of Correct Recognition
Hue-Pitch	34	6	85%
Saturation-Octave	25	15	62.5%
Intensity-Velocity	31	9	77.5%
Size-Gain	25	15	62.5%
Coordinate-Position	27	13	67.5%
Contents Label-Timbre	38	2	95%

And we apply CCS to blind persons in that experiment. Every blind person has profile Codes for identification in 〈Table 3〉. This result show that how much can understand with total color sense as feel of pictures and object coordinates by blind person.

표 3. 맹인에 의한 색의 인식 결과
Table 3. Results for Color Recognition by Blind Person

Profile Code	# of Samples	Similarity Recognition	Different Sense	Correct Total Recognition
01M	40	28	6	70%
02M	40	29	8	72.5%
03M	40	32	8	80%

V. Conclusion and Future Work

Interaction with image recognition can share and analyze before that understand mean of object in the human brain. Human perception is process generally express represent object like abstraction meaning. If we want transmit information to other acceptable person from some environment sense. It is require enormous semantic databases. In fact, semantic processes spend a lots of cost. Colored-hear Conversion System support fast transmission of vision data without any other semantic algorithms. The system first tracks object area using clustering algorithm; these object areas are then used to localize color description and feature vectors. The color factor and shape parameters corresponding to synesthesia space factors are recovered using Colored-hear Conversion System(CCS). Once the factors describing the synesthesia features are recovered, they are used to sending the auditory information. This approach is very efficient for recognize vision sensitivity which image objects and is able to feel that vision space robustly without eyes.

Colored-hear transformation for place recognition can get fast and detail perception. Also, It can be transmit information for sense at the same time. Thus, we were get a good result when applied this concepts to blind person's case of image recognition.

The framework suggested in this paper has some limitations. The system depends upon the object focus block, which currently breaks when the subjects are confused segmentation. The pattern recognition to find focus blocks can be further refined to track the important semantic area even when there are subjects with confused objects area. The synesthesia is a very important channel that combination signals related to the external state and a lot of effort is being devoted to unravel this relationship. Besides being used as a virtual interface or man-machine interface, this framework would hopefully be useful to a lot of these research efforts as well.

참고문헌

[1] Ward, J, B. Huckstep & E. Tsakanikos, "Sounds-colour synaesthesia: To what extent does it use cross-modal mechanisms common to us all?" Cortex Vol. 42, No. 2, pp. 264-280, 2006.

[2] Ramachandran, V.S. & E.M. Hubbard, "Synaesthesia: A window into perception, thought and language", Journal of consciousness Studies Vol. 8, No. 12, pp. 3-34, 2001.

[3] Karwoski, Theodore F., and Henry S.Odbert, "Color-Music, Psychological Monograpks, Vol. 50, No. 2, 1983.

[4] Baron-Cohen, S.; J. E. Harrison & L. H. Goldstein et al., "Coloured Speech Perception: Is Synesthesia What Happens When Modularity

Breaks Down?", Perception 22, pp. 419-426, PMID 8378132, 1993.

[5] Cytowic, R. E., "Synesthesia: A Union of the Senses", 2nd ed, Cambridge, MA: MIT Press, ISBN 0-262-03296-1, 2002.

[6] Cytowic, R. E., "The Man Who Tasted Shapes", New York, NY: Tarcher/Putnam, ISBN 0-262-53255-7, 2003.

[7] Smilek D., M. J. Dixon & C. Cudahy et al., "Synesthetic color experiences influence memory", Psychological Science, Vol. 13, No. 6, pp. 548-552, 2002.

[8] Sagiv N., J. Heer & L. C. Robertson, "Does binding of synesthetic color to the evoking grapheme require attention?", Cortex Vol. 42, No. 2, pp. 232-242, PMID 16683497, 2006.

[9] Robertson, L. C. & N. Sagiv, "Synesthesia: Perspectives from Cognitive Neuroscience", Oxford, UK: Oxford University Press., ISBN 0-19-516623-X, 2005.

저 자 소 개

신 성 윤

2003년 2월 군산대학교 컴퓨터과학과 이학박사

2006년~현재 군산대학교 컴퓨터정보과학과 교수

〈관심분야〉 비디오 인덱싱, 비디오 요약, 멀티미디어

문 형 윤

1997년 명지대학교 산업기술대학원 컴퓨터공학과 졸업(석사)

2007년 육군본부 전산장교

1997~현재 LG CNS 공공사업본부 차장

〈관심분야〉 영상처리, 컴퓨터비전

표 성 배

2004년 2월 숭실대학교 컴퓨터학부 공학박사

1992년~현재 인덕대학 소프트웨어개발과 교수

〈관심분야〉 멀티미디어, 비디오 색인화, 비디오 요약