HMM(Hidden Markov Model)을 이용한 Finished Paint의 배합비율 예측에 관한 연구

조용훈*, 이성근, 안대명, 황규석
부산대학교 화학공학과
(ssauza_md@yahoo.co.kr*)

Formulation ratio Prediction of Finished Paint using Hidden Markov Model

Yong Hoon Cho*, Sung Gun Lee, Dae Myung An, Kyu Suk Hwang
Dept. of Chemical Engineering, Pusan National University
(ssauza_md@yahoo.co.kr*)

INTRODUTION

색은 환경, 건축, 패션, 화장품, 멀티미디어 등 다양한 분야에서 사용되고 있으며 국제시장의 산업경쟁력에 있어 그 중요성은 더욱 커지고 있다. 우리나라에서도 다양한 색상 제품을 제공하기 위해 보다 정확한 system을 구축하고자 하는 노력이 계속 되고 있지만 이런 system들은 수학적인 모델링이 근거가 되어야 하기 때문에 프로세스의 동 특성에 대한 정확한 지식이 필수적이며 제어대상이 너무 복잡하거나 비 선형성이 강한 경우에는 만족스러운 결과를 얻기가 힘들다. 때문에 현재 사용되는 CCM(Computer Color Matching) 및 database는 거의 수입에 의존하고 있는 실정이다. 일반적으로 도료산업에서 사용되는 CCM은 빛의 산란계수와 흡수계수를 다룬 Kubelka-Munk 방정식을 사용하고 있다. 이는 매우 복잡하여 작업상에 어려움이 있을 뿐만 아니라, metallic 또는 pearlescent color 에는 적용하지 못하며 3가지 이상의 혼합 색은 잘 맞지 않는다는 단점을 가지고 있다.

본 연구에서는 Hidden Markov Model을 이용하여 색 혼합의 복잡하고 강한 비 선형성을 가지는 다양한 문제를 해결할 것이다. 또한 복잡한 수학식이 없고 추가 색상에 의한 모델링이 간편하여 작업자가 편하게 사용할 수 있다. 본 연구에서는 White, Blue, Green 3가지색의 배합비율 data로 각각의 HMM Parameter를 구하고 이 Parameter를 이용하여 임의의 색상에 대한 배합비율을 예측 및 평가하였다.

PROBLEM FORMULATION

HMM(Hidden Markov Model) : 통계적 과정을 갖는 유한 상태 구조로 시간에 따른 상태간의 변화를 결정하는 천이 확률과 각 상태에서의 출력확률로 구성된다.

λ=(Π, A, B) → Model Parameters

Π : 초기상태 확률분포의 집합

Π={π_1, π_2, π_3, …, π_N}

πi=P(s=s_i | λ), 1≤ i ≤N , π_i는 초기상태가 s_i 인 확률, N : 상태(state) 수

A : 상태천이 확률분포의 집합

A={a_{ij}}, 1≤i, j≤N 1≤t≤T

a_{ij}=P(st+1=sj | st=si), a_{ij}는 시간 t에서 상태 s_i로부터 시간 t+1에서 s_j로의 천이확률

B : 상태에서의 관측확률의 집합

B=b_j(k), 1≤j≤N, 1≤t≤T, 1≤k≤M

$b_j(k)=P(o_t=v_t \mid s_t=s_j)$, $b_j(k)$는 시간 t에서 상태 j에서의 심볼 k의 관측확률

PROBLEM SOLUTIONS

HMM을 사용할 때에는 1)시스템 상태의 가장 최적의 상태열을 찾기위한 해석 문제(Decoding problem)와 2)관측열에 대하여 확률을 최대화하는 모델 변수를 예측하는 문제(Estimation problem)가 있다.

이상의 기본적인 문제점에 대한 해결 방안으로 다음 연구에서는 전향-후향 알고리즘, viterbi 알고리즘, Baum-Welch 재추정 알고리즘을 사용할 것이다.

A. 전향 알고리즘(forward algorithm)과 후향 알고리즘(backward algorithm)

전향변수 α_t는 초기상태에서 시작하여 o_1, o_2, …, o_t를 생성하면서 상태 i에 도달하는 확률로 정의한다. 이 전향확률은 다음처럼 재귀적으로 계산하는 것이 가능하다.

$$\alpha_1(i) = \pi_i b_i(o_1), \qquad 1 \le i \le N$$

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(o_{t+1}), \qquad 1 \le t \le T-1$$

후향변수 β_t는 최종상태에서 처음으로 o_T, …,o_{t+2},o_{t+1}을 생성하면서 상태 i에 도달하는 확률로 정의하면

$$\beta_T(i) = 1, \qquad 1 \le i \le N$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t-1}) \beta_{t+1}(j), \qquad t = T-1, T-2, ..., 1$$

B. 비터비(Viterbi Decoding)

viterbi 알고리즘은 최적 원리에 입각한 동적 프로그래밍 기법의 하나이다. 이 알고리즘은 임의 상태에 이르는 경로 비용 또는 확률을 계산할 때, 이전 상태까지의 비용과 이전 상태들로부터 현재 상태로의 전이 비용을 곱하는 방식으로 순환계산하는 기법이다.

초기화 $\delta_1(i) = \pi_i b_i(O_1)$, $\psi_1(i) = 0$

순환계산 $\delta_t(j) = \max_i [\delta_{t-1}(i) a_{ij}] b_j(O_t)$, $\psi_t(j) = \arg\ \max_i [\delta_{t-1}(i) a_{ij}]$

종료 $P^* = \max_{1\le i\le N} [\delta_T(s)]$, $s^*{}_T = \arg\max_{1\le i\le N} [\delta_T(s)]$

역추적 $t = T-1, \cdots, 1$, $s^*{}_t = \psi_{t+1}(s^*{}_{t+1})$

C. 재추정(reestimation)

전방향 확률과 역방향 확률을 이용하여 학습패전에 대하여 모델의 특정상태 사이에서 발생한 전이 횟수의 기대값 $\xi_t(i,j)$ 와 특정 상태로부터 발생한 전이 횟수의 기대값 γ_t (i)를 구한다.

$$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i.j)$$

$$\xi_t(i,j) = P(s_t = i,\ s_{t+1} = j | o, \lambda)$$

$$= \frac{\alpha_t a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{P(o|\lambda)} = \frac{\alpha_t a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}$$

D. 변경(update)

주어진 학습 패턴에 대해 전이 횟수의 기대값 $\xi_t(i,j)$ 와 γ_t (i)를 이용하여 확률 parameter $\overline{\lambda} = (\overline{\Pi}, \overline{A}, \overline{B})$ 는 다음과 같은 새로운 값으로 된다.

$$\overline{\Pi}_1 = \gamma_1(i) \qquad \overline{a_{ij}} = \frac{\sum_{t=1}^{T-1} \gamma_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \qquad \overline{b_j}(k) = \frac{\sum_{t=1,\ s.t.\ o_t = v_k}^{N} \gamma_t(i)}{\sum_{t=1}^{N} \gamma_t(j)}$$

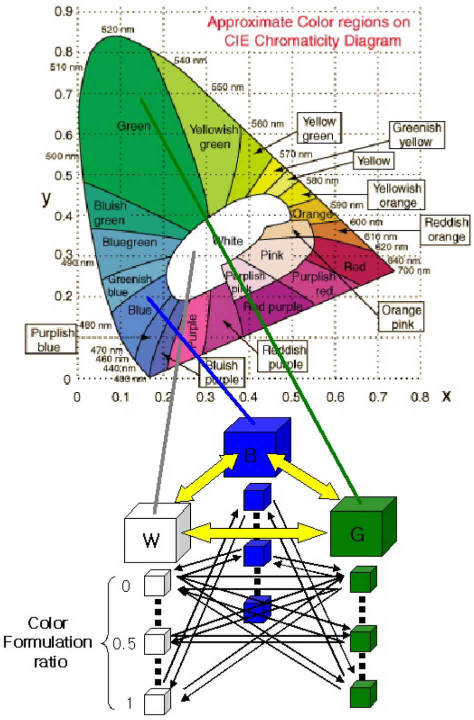
위의 parameter로 결정된 최적경로의 값이 더 이상 증가 하지 않을 때 까지 parameter 값을 계속 변경 시킨다.

SIMULATION

본 연구의 계산은 MatLab을 사용하였으며 프로그램에서의 상태열은 각 컬러의 매칭비율, 관측열은 CIE표색계의 XYZ값을 Lab로 변경한 값을 사용하였다.

White, Blue, Red의 배합비율을 0.1~0.8까지로 각각 조절하여 거기에 따른 parameter를 아래와

같이 계산하여 학습 시킨 후 5가지의 sample color를 test하여 그 성능을 확인하였다.



<Overview of color formulation ratio>

i	0.0035	0.0040	0.6488	0.3305	0.0068	0.0061	0.0001	0.0001
t	0.0041	0.0041	0.0042	0.0043	0.0044	0.0062	0.3706	0.7797
	0.0044	0.0044	0.0044	0.0045	0.0046	0.0065	0.4210	0.0009
	0.6613	0.6566	0.6642	0.6721	0.6780	0.9858	0.1363	0.1435
	0.3142	0.3211	0.3156	0.3132	0.3130	0.0014	0.0692	0.0729
	0.0060	0.0059	0.0059	0.0059	0.0000	0.0000	0.0014	0.0015
	0.0058	0.0057	0.0057	0.0000	0.0000	0.0000	0.0014	0.0014
	0.0021	0.0023	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001
	0.0021	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0001
o	0.6327	0.6454	0.6603	0.6802	0.7061	0.7063	0.5101	0.5107

<L:56.5553 a:-35.3275 b:16.4367의 최종parameter>

L	a	b	최대값(p)
30.4000	-38.5802	21.5579	2.3387e-004
46.9745	-38.5558	18.8950	7.0126e-006
56.5553	-35.3275	16.4367	2.0668e-005
67.9911	-28.1971	12.1185	9.3968e-006
80.6163	-17.3115	6.7075	1.0592e-005

<sample color의 최대값(p)>

CONCLUSION

이 연구에서 우리는 2가지 주요한 점에 집중하였다. 관측된 실험 결과를 기초로 HMM parameter들을 유도하였고, 기존의 데이터를 이용하여 HMM의 거동을 분석하였다. 또한 도료의 배합비율을 달리하여 색의 변화를 비교하고, 학습의 속도와 색차를 줄일 수 있는 모델을 연구하였다. 그 방법은 각각의 배합비율을 고정된 확률적 과정으로 모델링하여 결합시키는 것이다. 이것으로 우리는 하나의 색에 대하여 고유한 parameter를 얻을 수 가있고 더욱 정확한 Color recipe prediction과 Color matching이 가능하다.

REFERENCE

1. Warren J. Ewens, Gregory R. Grant, "Statistical Methods in Bioinformatics:An Introduction". 129-348
2. 김상운 "식별 알고리즘을 중심으로 한 패턴인식 입문" 137-152
3. Hans G.Vőlz "Industrial Color Testing"
4. 김삼수, 박성수 "디지털 색상의 원리와 응용"