

Traverten Plaka Taşlarda Sınıfların K-ortalamlar ve Bulanık C-ortalamlar Kümeleme Yöntemleri ile Belirlenmesi

Melike ŞİŞECİ^{1*}, Bayram CETİŞLİ¹

¹ Süleyman Demirel Üniversitesi, Bilgisayar Mühendisliği Bölümü / ISPARTA

Alınış Tarihi:20.11.2011, Kabul Tarihi:23.10.2012

Özet: Doğal taşlarla yapılan yüzey kaplamalarında, yüzeylerin homojen olarak görünmesi estetik açıdan önemlidir. Bu noktada gerek müşteri talepleri gerek taş üreticileri tarafından belirlenmiş sınıflar bulunmaktadır. Üretim sırasında taş plakaların sınıflandırılması (seleksiyonu) genellikle elle yapıldığından sınıflarda bir standardın tutturulması mümkün olmamaktadır. Ayrıca önceden belirlenen sınıf standartlarına uygun olmayan taşların gelmesi durumunda da en yakın sınıf tespit edilerek plakalar o sınıflara ayrılmaktadır. Bu noktada kişilerin farklı görselliklerinden bağımsız olarak objektif bir sınıflandırmanın yapılması gerekmektedir. Bu çalışmada ise öğreticisiz öğrenme modellerinden biri olan K-ortalamlar ve bulanık C-ortalamlar yöntemleri birbirleriyle kıyaslanarak traverten plakalar için yeni bir sınıf tanımlaması öngörülmektedir. Yöntemlerde, hem renk hem de homojenlik analizleri yapılarak travertenler için en uygun öznelilikler belirlenmiştir. Yapılan deneysel çalışmalar, önerilen yeni sınıflandırma biçimleri ile traverten sınıflarının ayrımı daha belirgin olduğunu ve sınıf içi dağılımlarının göze daha hoş geldiğini göstermektedir.

Anahtar Kelimeler: K-ortalamlar kümeleme yöntemi, Bulanık C-ortalamlar, Homojenlik analizi, Renk analizi, Traverten sınıflandırma, Görüntü işleme.

Determination of Travertine Plate Stone Classes with K-means and Fuzzy C-means Clustering Methods

Abstract: It is aesthetically important that surfaces appear as homogeneous in the surface coverings made of natural stones. At this point, there are classes determined by both the customers' demands and the manufacturers of stone. During the production, since the classification of stone plates (selection) is generally done manually, it is not possible to reach a standard. In addition, the closest class is determined and plates are divided into these classes in case stones are non-compliance with pre-determined class standards. At this point, an objective classification should be done being independent from people's different visuality. In this study, K-means which is one of the learning models without a teacher and fuzzy C-means methods are compared with each other, new class definition is foreseen for travertine slabs. In the methods, analyzing both color and homogeneity, the most suitable qualities have been determined for travertine's. Experimental studies show that the discrimination between the proposed new classification forms and travertine classes is more obvious and the distribution of them in classes seems more pleasant.

Key Words: K-means clustering methods, Fuzzy C-means, Homogeneity analysis, Color analysis, Travertine classification, Image Processing.

Giriş

Son yıllarda yapı sektöründeki gelişmeler sonucunda doğal taş talebi oldukça artmaktadır. Dünyanın en zengin doğal taş oluşumlarının bulunduğu Alp kuşağında yer alan Türkiye, çok çeşitli ve büyük miktarda doğal taşlar rezervine sahiptir (Eraslan vd., 2008). Türkiye, bu kaynaklara ilaveten gelişmekte olan sanayisi ve üretimde kullandığı teknoloji ile dünyanın en önemli doğal taş üreticileri arasında yer almaktadır. Türkiye'de doğal taş zenginliği, binlerce yıl önce yapılan tarihi eserlerde de görülmektedir. Taşların tarih boyunca kullanımı, turizm açısından önem arz etmektedir. Dünyadaki doğal taş talebi, bu konuda büyük bir zenginliğe sahip Türkiye'nin doğal taş üretiminde ön sıralarda yer almasını sağlamıştır.

Üretim için teknolojik gelişmelerin önemi büyüktür. Taşların madenlerden çıkarılması ve işlenmesi için büyük enerji kaynaklarına ve iş makinelerine gerek vardır. Türkiye'de bu konuda çok iyi işletmeler bulunmaktadır. Asıl sorun ürünün sınıflandırılması aşamasında oluşan kişilere bağlı hatalardır. Bu noktada hataları en aza indirmek için otomatik makine görmesi ile çalışan sistemlerin kurulması gerekmektedir. Otomatik görme sistemleri, seleksiyon işçilerinin yaptıkları ve yapabilecekleri hataları ortadan kaldırmakta ve

seleksiyon işlemlerini hızlandırmaktadır (Metlek vd., 2009).

Doğal taş seçiminde en önemli unsur renk uyumudur. Bunun yanı sıra yüzeyde farklı malzemelerin varlığından dolayı farklı doku ve renkte büyüklü küçüklü bölgeler bulunmaktadır. Bu bölgelerin varlığı, taşın estetik güzelliğini etkilemektedir. Benzer taş plakların bir araya gelmesi durumunda, gözde bir akışkanlık oluşturması beklenir. Böylesine yüzeylerde göz, hemen aykırı olanı bulur ve ona takılır. İnsanın tanıma sisteminde, daha önce gördüğü bir varlığı, yeni gördüğü ortamlarda da arama, benzerlik ve ilişki kurma gibi sürekli yaptığı durumlar bulunmaktadır. Örneğin bulutlara baktığında bulutları bir insan yüzüne, bir varlığa benzetmesi bu sistemin bir sonucudur. Özellikle zıtlığı çok fazla olan yüzeylerde bu arayışa daha çabuk girmektedir. Sanatsal ortamlarda benzerliği yüksek olan taşlar kullanmak yerine çağrışım yapan zıtlığı fazla olan taşlar kullanmak daha mantıklıdır. Bu noktada yüzey kaplamalarında, benzerliğin önemi kadar zıtlığın da önemi olmalıdır.

Homojenite (HM) kavramında benzerlik özelliklerinin varlığı ve dereceleri HM ile doğru orantılıdır. Taşın homojenliği; taştaki yabancı elemanlar, taşın renk ve

* ustatmel@hotmail.com

kimyasal bileşimleri parametrelerine bağlıdır. Taşın renklenmesinde, taşın içerisine giren yabancı elemanlar önemlidir. Metamorfik taşlarda metamorfizma derecesi de, özellikle iç yansıma ve aynı mineral grubu kristallerin farklı sistemlerde gelişimine bağlı olarak taşın renklenmesini etkilemiştir. Taş içerisine giren her farklı mineral, taşın kimyasal içeriğini de etkiler (Karaca ve Öztank, 2003).

Türkiye sanayisinde taş işleme endüstrisinin ekonomik değeri oldukça büyüktür. Yüzey kaplamasında kullanılan taşların müşteri isteği ve çıkan taşın çeşitliliğine göre bazı sınıflar oluşturulmuştur. Üreticiler bu sınıflandırmaya göre ürettikleri taşları tasniflemektedirler. Ancak çıkarılan taşların farklı madenlerden olması, taş damarlarının farklılık göstermesi nedeniyle yanlış sınıflandırmalar oldukça fazladır. Ayrıca sınıfların belirlenmesinde belirli bir standart oluşturulamamıştır. Buna hangi taşın hangi sınıfa ayrılacağını tam olarak bilemeyen kalite kontrol işçilerinin hataları da eklenebilir. Bu noktada temel sınıfları belirlemek ve sınıf özelliklerinin tespit edilebilmesi adına yapılabilecek en iyi yöntem taşları kümelemektir. Kümeleme yöntemleri önyargısız biçimde hiçbir önceden belirlenmiş sınıf etiketlerini kullanmadan kendisi kümeleri oluşturduğu için yeni sınıfların oluşturulması adına bir katkı sağlayacağı kesindir. Bu çalışmada benzerlik ve zıtlıkları, renk tonları, farklı renkte bölge alanları dikkate alan bir kümeleme ile traverten sınıfları belirlenmiş ve oluşan kümelerdeki traverten plaka resimleri bir arada verilerek göze hoş gelen bir yüzey oluşumu sağlanmaya çalışılmıştır. Ayrıca üretici firmanın yapmış olduğu sınıflandırma sonuçları ile bizim önerdiğimiz sınıflandırma sonuçları karşılaştırılarak yapılan sınıflandırma hataları ortaya konmuştur.

Konu ile ilgili çalışmalar incelendiğinde daha çok mermer, granit gibi taşların sınıflandırılmasının ve doku analizlerinin yapıldığı görülmektedir. Martinez-Alajarin vd., mermer yüzeyleri üzerinde homojenlik, zıtlık, enerji, ortalama ve standart sapma gibi istatistiksel öznitelikleri incelemiş ve farklı doğrusal renk uzaylarındaki öznitelikleri sınıflandırmada kullanmıştır (Martinez-Alajarin vd., 2005). Selver ve Akay, toplam ve fark histogram özniteliklerini kullanarak otomatik mermer sınıflandırmayı yapmışlardır (Selver ve Akay, 2009). Benavente ve Pina mermer dokularını matematiksel morfoloji ile bölütlemiş ve sınıflandırmışlardır (Benavente ve Pina, 2009). Lopez vd., ise granit taşları spektral özelliklerine göre sınıflandırmışlardır (López vd., 2010).

Özniteliklerin belirlenmesinde ticari firmalar sadece rengi ve renklerin istatistiksel özniteliklerini ele almışlardır. Bu noktada doku analizlerinin yapılmadığı görülmektedir. Doku analizi, tıp, tekstil, uzaktan algılama gibi alanlarda sıklıkla kullanılmaktadır. Yüzeylerin ayrıştırılması, bölütleme, farklı olanı ortaya çıkarma gibi amaçlar için doku analizi yöntemlerine başvurulmaktadır. Bu noktada Fourier, dalgacık gibi yöntemlerle dokunun farklı frekansları belirlenmekte, spektral, istatistik ve yapısal özellikleri ortaya konmaktadır (Gonzalez ve Woods, 2008).

Sınıfların belirlenen özniteliklere göre yeniden belirlenmesi aşamasında ise daha çok bir yapay zekâ uygulaması olan kümeleme yöntemleri öne çıkmaktadır. Yapay zekâ yöntemlerinde sistemlerin var olan problemleri çözebilmesi için tasarlanan sistemlerin eğitilmesi gerekmektedir. Bu noktada iki tip eğitim yöntemi öne çıkmaktadır. Bunlar öğreticili ve öğreticisiz öğrenmedir (Jang vd., 1997). Öğreticili öğrenmede; eğitim setindeki örneklerin her birinin hedef değerleri bellidir ve sistem bu hedeflere ulaşmak için kendi parametrelerini uyarlar. Hata belli bir değer altına indiğinde ise öğrenme tamamlanır. Sistem oluşturmada en çok kullanılan yöntem budur. Yapay sinir ağları, sinir-bulanık sistemler ve destek vektör makineleri öğreticili öğrenmeyi en çok kullanan yöntemlerdir. Bunun yanı sıra, hedeflerin belli olmadığı öğrenmede, sistem örnekler üzerinden kendi kendine öğrenebilmektedir. Kendi kendine öğrenen yöntemlerin birçoğu kümeleme yöntemi olmasına rağmen, Kohonen ve Hopfield gibi sinir ağları da öğreticisiz öğrenme yöntemleri arasında sayılabilir (Kohonen,1977; Hopfield, 1982).

Kümeleme yöntemleri, eğitim setindeki sınıfları belli olmayan örnekleri benzerliklerine göre gruplara ayırırlar. Kümeleme bittikten sonra belli sayıda oluşan gruplar etiketlenerek grupların merkezleri ve büyüklükleri belirlenir. Kümeleme yöntemleri bazen başka öğreticili sistemlerin eğitimde kullanacakları başlangıç değerlerini belirlemede de kullanılabilir. Ancak daha çok veri bölütleme, veri madenciliği, örnekleme gibi alanlarda kullanılmaktadır. Kümeleme yöntemleri hiyerarşik olan ve olmayan biçiminde ikiye ayrılabilir. Hiyerarşik kümeler istatistiksel yöntemler olup en bilinenleri en yakın komşu ve en uzak komşu algoritmalarıdır (Jang vd.,1997; Özkan, 2008). Hiyerarşik olmayan yöntemler genellikle tekrarlamalı yöntemler olup K-ortalamlar, bulanık C-ortalamlar, çıkarıcı, dağ algoritmaları en bilinen yöntemlerdir (Jang vd.,1997). Bunların arasında K-ortalamlar tutarlılığı ve hızı açısından öne çıkan ve bu yüzden en çok kullanılan kümeleme yöntemidir.

Materyal ve metot kısmında homojenlik, benzerlik, zıtlık gibi doku analizi kavramları, taş resimlerinden çıkarılan öznitelikler ile hiyerarşik olmayan kümeleme yöntemlerinden K-ortalamlar ve bulanık C-ortalamlar verilmektedir. Bulgular kısmında K-ortalamlar ve C-ortalamlar yöntemleri kullanılarak yapılan deneysel çalışmalar şekil ve tablolarla açıklanmıştır. Son olarak tartışma ve sonuç kısmında yapılan çalışmaların değerlendirilmesi yapılmıştır.

Materyal ve Metot

Öznitelik çıkarma ve öznitelik seçimi

Örüntü tanımanın temel basamakları, öznitelik çıkarma, gürültü temizleme ve normalleştirme, öznitelik seçimi ve sınıflandırmadır. Sınıflandırmada kullanılabilecek değişkenler, örüntüyü diğer örüntülerden ayırt edebilecek yapısal ve istatistiksel özelliklerdir. Bu çalışmada traverten taşların görüntüleri ele alındığından, görüntü işlemede kullanılan öznitelikler ele alınmaktadır. İşlenen

taşların yüzeysel olması, renk ve doku özneliklerinin çıkarılmasını gerektirir.

Doku analizi

Doku analizinde kullanılan yöntemler istatistiksel yöntemlerdir. Bunlardan en bilineni eş-oluşum (co-occurrence) matrisleridir. Eş-oluşum matrisleri bir resimde ardışık biçimde gelen piksel değerlerinin ve sayılarının tespitidir (Castleman, 1996). Eş-oluşum matrisi ile entropi, atalet, moment, enerji, zıtlık, HM, ilinti, ortalama gibi istatistiksel değerleri hesaplamak mümkündür (Castleman, 1996; Gonzalez ve Woods,

2008; Martinez-Alarajin vd., 2005). Eş-oluşum matrisinin elde edilmesi ile ilgili basit bir örnek Şekil 1'de verilmektedir.

Gri seviye eş-oluşum matrisi (GSEOM) G , kullanılarak p_{ij} olasılık değerleri hesaplanır, burada i ve j ardışık gelen piksel değerlerini göstermektedir. Buna göre çalışmada kullanılan zıtlık ve HM ifadeleri (Gonzalez ve Woods, 2008):

$$Zıtlık = \sum_{i=1}^K \sum_{j=1}^K (i-j)^2 p_{ij}, \quad (1)$$

$$Homojenite = \sum_{i=1}^K \sum_{j=1}^K \frac{p_{ij}}{1+|i-j|}. \quad (2)$$



a) Özgün gri seviye resim

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

b) Resimin piksel değerleri

Referans Piksel Değeri	Komşu Piksel Değeri			
	0	1	2	3
0	(0,0)	(0,1)	(0,2)	(0,3)
1	(1,0)	(1,1)	(1,2)	(1,3)
2	(2,0)	(2,1)	(2,2)	(2,3)
3	(3,0)	(3,1)	(3,2)	(3,3)

c) Piksel komşuluk matrisi

	0	1	2	3
0	2	2	1	0
1	0	2	0	0
2	0	0	3	1
3	0	0	0	1

d) Eş-oluşum matrisi

Şekil 1. Eş-oluşum matrisinin elde edilmesi

Renk analizi

Tek başına doku analizi, traverten sınıf belirlemede yetersiz kalmaktadır. Bu noktada renk analizleri önem kazanmaktadır. Ancak sadece Red-Green-Blue (RGB) (Kırmızı-Yeşil-Mavi) renk uzayı ile bazı örnekleri ayırmak mümkün olmamaktadır. Bu nedenle farklı doğrusal ve doğrusal olmayan renk uzayı dönüşümleri de bu çalışmada kullanılmıştır. Doğrusal olanlar YIQ (Brightness In-phase Quadrature) ve YCbCr (Brightness Chroma-Blue Chroma-Red) dönüşümleridir (Chaves-González, 2010):

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & 0.274 & 0.322 \\ 0.211 & 0.253 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (3)$$

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4)$$

Doğrusal olmayan renk uzayları ise HSV (Hue Saturation Value), CIE- L^*c^*h (Lightness Chroma Hue) ve DAC (Digital to Analog Converter) dönüşümleridir. HSV dönüşümü (Chaves-González, 2010):

$R, G, B \in [0,1]$, $h \in [0, 360]$, $M = \max\{R, G, B\}$, $m = \min\{R, G, B\}$ olmak üzere;

$$h = \begin{cases} 0, & \text{Eğer } M = m, \\ \left(60^\circ \frac{G-B}{M-m} + 360^\circ\right) \bmod 360^\circ, & \text{Eğer } M = R \\ 60^\circ \frac{B-R}{M-m} + 120^\circ, & \text{Eğer } M = G \\ 60^\circ \frac{R-G}{M-m} + 240^\circ, & \text{Eğer } M = B \end{cases} \quad (5)$$

$$s = \begin{cases} 0, & \text{Eğer } M = 0 \\ \frac{M-m}{M} = 1 - \frac{m}{M}, & \text{Eğer } M \neq 0 \end{cases}$$

$$v = M$$

CIE- $L^*a^*b^*$ dönüşümünden üretilen CIE- L^*c^*h renk uzayının ifadesi (Osareh vd., 2002);

$$L^* = 116 \sqrt[3]{\frac{Y}{Y_n}} - 16$$

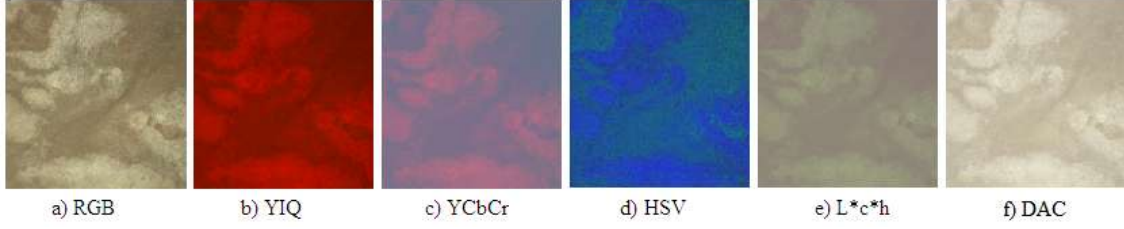
$$c^* = \{(a)^2 + (b)^2\}^{0.5}. \quad (6)$$

$$h = \arctan\left(\frac{b}{a}\right)$$

DAC sayısal sinyalleri analog sinyallere dönüştürür. RGB renk uzayıyla 8-bit DAC arasındaki dönüşüm aşağıda verilmektedir (Viénot vd., 1999);

$$\begin{aligned} I_d &= 255 \times R_d^{\frac{1}{2.2}} \\ J_d &= 255 \times G_d^{\frac{1}{2.2}} \\ K_d &= 255 \times B_d^{\frac{1}{2.2}} \end{aligned}, \quad (7)$$

$$0 \leq R_d \leq 1, 0 \leq G_d \leq 1, 0 \leq B_d \leq 1$$



Şekil 2. Renk uzayları

Pearson ilinti ölçütü

Oluşturulan öznelik sayısı çok fazla olduğundan bunların işe yaramayanları elenerek, en iyi öznelikler sınıf belirlenmesinde kullanılmalıdır. Bu notada kullanılan birçok öznelik seçimi ya da boyut azaltma yöntemleri kullanılabilir. Bunlardan en bilineni ve hesaplaması en kolay olan yöntem Pearson ilinti ölçütüdür.

Genellikle giriş değişkenlerinin çıkış değişkenleriyle ilişkili olması gerektiği beklenir. İlişkileri belirlemek için aşağıdaki ifadeler kullanılır. N ve D 'nin sırasıyla örnek sayısı ve öznelik sayısını gösterebilir. $\mathbf{X} \in \mathbb{R}^{N \times D}$ giriş matrisini ve \mathbf{X}_i bu matristeki i . özneliği ifade etsin ve \mathbf{Y} ise giriş değerlerine göre belirlenen sınıf etiketlerini gösterebilir. Buna göre Pearson ilinti katsayısı şu şekilde ifade edilir:

$$R(i) = \frac{\sum(\mathbf{X}_i \mathbf{Y})}{\sigma_{\mathbf{X}_i} \sigma_{\mathbf{Y}}}, i = 1, 2, \dots, D. \quad (8)$$

Eşitlik (8)'de \sum özdeşinti matrisini, σ standart sapmayı göstermektedir (Theodoridis ve Kotroumbas, 1999). $R(i)$ genellikle çıkış değerleriyle ilişki yeteneklerine göre öznelikleri sıralamada kullanılır.

Hiyerarşik olmayan kümeleme yöntemleri

Hiyerarşik olmayan sınıfa giren kümeleme algoritmaları veriyi doğrudan kümeleyen algoritmalar (Taşkın ve Emel, 2010). Kümeleme analizi $X \in \mathbb{R}^{N \times d}$ veri tabanında yer alan ve doğal sınıfları kesin olarak verilmeyen örnekleri belli bir uzaklık kavramıyla gruplandırarak ve alt kümeleri oluşturan yöntemlerdir (Özdamar, 2004). Genellikle rastgele başlangıç değerleri ile sürekli küme içi dağılımları en küçük yapan ötelemeli algoritmalarla kümeler oluşturulur.

K-ortalamar kümelemesi

En eski kümeleme algoritmalarından olan K-ortalamar, 1967 yılında MacQueen tarafından geliştirilmiştir (MacQueen, 1967). En yaygın kullanılan öğreticisiz öğrenme yöntemlerinden birisi olan K-ortalamarın

Burada I_d kırmızı (R_d), J_d yeşil (G_d), K_d ise mavi (B_d) rengin Deutan değiştirme ile elde edilen analog sinyal değerleridir.

Şekil 2, örnek bir travertenin beş renk uzayında elde edilen görüntülerini göstermektedir.

atama mekanizması, her verinin sadece bir kümeye ait olabilmesine izin verir. Bu nedenle, keskin bir kümeleme algoritmasıdır. Merkez noktanın kümeyi temsil etmesi ana fikrine dayalı bir yöntemdir (Han ve Kamber, 2006). K-ortalamar algoritmasında her bir nesnenin merkez noktalara uzaklığını hesaplamak için en çok kullanılan iki farklı yöntem aşağıda açıklanmaktadır (MacQueen, 1967):

Öklid uzaklığı

$1 \times d$ boyutlu iki örnek arasındaki mutlak uzaklığı gösterir. Bu yönüyle Pisagor teoreminin bir uygulamasıdır (Periavi ve Kheibari, 2008). Sonuç olarak Öklid uzaklığı $O(i, j)$:

$$O(i, j) = \sqrt{\sum_{k=1}^d (x_{ik} - x_{jk})^2}. \quad (9)$$

Burada i ve j iki ayrı örneğin indislerini göstermektedir.

Manhattan uzaklığı

Manhattan uzaklığı örnekler arasındaki mutlak uzaklıktır. Manhattan uzaklığının ifadesi şöyledir. (Höpler vd., 1999).

$$M(i, j) = \sum_{k=1}^d (|x_{ik} - x_{jk}|). \quad (10)$$

Verilen uzaklık ölçütlerinin dışında Minkowski, Pearson gibi başka uzaklık ölçütleri literatürde tanımlanmış ve kullanılmıştır. Ancak en yaygın olarak kullanılan ve bilineni Öklid uzaklığıdır.

K-ortalamar algoritması N adet \mathbf{x}_j örneklerini c grup sayısına bölütler ve her grup için uzaklık ölçümü olarak tanımlanan amaç fonksiyonunu küçültme yoluyla küme merkezlerini bulmaya çalışır. Eğer uzaklık ölçütü Öklid uzaklığı seçilirse herhangi bir gruba dâhil edilen \mathbf{x}_k örneklerinin \mathbf{c}_i küme merkezine olan uzaklıkları hesaplanarak toplam uzaklık maliyeti şu şekilde tanımlanır:

$$J = \sum_{i=1}^s J_i = \sum_{i=1}^s (\sum_{k, \mathbf{x}_k \in G_i} \|\mathbf{x}_k - \mathbf{c}_i\|^2), \quad (11)$$

Burada J_i i . grubun amaç fonksiyonudur. J_i , G_i kümesinin geometrik özelliklerine bağlıdır ve merkezi c_i 'dir (Jang vd., 1997).

Parçalanmış gruplar genellikle $c \times n$ boyutundaki U üyelik matrisi ile tanımlanır. Eğer x_j verisi i . grup altında kalırsa u_{ij} değeri 1'dir, değilse 0'dır. Bu durum Eşitlik (12)'de gösterilmektedir.

$$u_{ij} = \begin{cases} 1, & \text{Eğer } \|x_k - c_i\|^2 \leq \|x_k - c_k\|^2, \text{ her bir } k \neq i \text{ için} \\ 0, & \text{Diğer durumlarda} \end{cases} \quad (12)$$

Bütün veriler bir gruba dâhil edildiğinden dolayı U matrisi şu özellikleri sağlar.

$$\sum_{i=1}^s u_{ij} = 1, \forall j = 1, \dots, n \text{ ve } \sum_{i=1}^s \sum_{j=0}^n u_{ij} = n \quad (13)$$

Eğer u_{ij} sabitlenirse en iyi küme merkezi c_i Eşitlik (11)'i küçültür ve c_i 'nin yeni değeri:

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} x_k, \quad (14)$$

Burada $|G_i|$ kümesine ait eleman sayısını gösterir.

Ötelemeli bir yöntem olan K-ortalamalar algoritması verilen sayıdaki kümelerin merkezlerini ve aitlik matrisini her adımda yeniler. Amaç fonksiyonu yeteri kadar küçültüldüğünde ya da verilerin aitliği değişmediğinde algoritma sonlandırılır.

Bulanık C-ortalamalar kümelemesi

Bulanık C-ortalamalar (Fuzzy C-means Clustering) algoritması, bulanık bölünmeli kümeleme tekniklerinden en iyi bilinen ve yaygın kullanılan yöntemidir. Bulanık C-ortalamalar algoritması 1973 yılında Dunn tarafından ortaya atılmış ve 1981'de Bezdek tarafından geliştirilmiştir (Dunn, 1973; Işık ve Çamurcu, 2007). Bulanık C-ortalamalar algoritması amaç fonksiyonunu küçültmeyi hedefler. Bulanık C-ortalamalar metodu, nesnelerin iki veya daha fazla kümeye ait olabilmesine izin verir. Bulanık mantık prensibi gereği her veri, kümelerin her birine $[0,1]$ arasında değişen birer üyelik değeri ile aittir. Bir verinin tüm sınıflara olan üyelik değerleri toplamı Bayes teoremindeki gibi 1 olmalıdır. Nesne hangi küme merkezine yakın ise o kümeye ait olma üyeliği diğer kümelerle ait olma üyeliğinden daha büyük olacaktır. Amaç fonksiyonun belirlenen minimum değere yakınsamasıyla kümeleme işlemi tamamlanır (Işık ve Çamurcu, 2007). Ötelemeli bir algoritmadır ve en küçük kareler yönteminin genellemesi olan aşağıdaki amaç fonksiyonunu küçültmeye çalışır.

$$J_m = \sum_{i=1}^N \sum_{j=1}^S u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty \quad (15)$$

Eşitlik (15)'de m ağırlık değeridir.

U üyelik matrisi rastgele atanarak algoritma başlatılır. İkinci adımda ise merkez vektörleri hesaplanır. Merkezler aşağıdaki ifade ile hesaplanır (Işık ve Çamurcu, 2007).

$$c_j = \frac{\sum_{i=1}^N u_{ij}^m x_i}{\sum_{i=1}^N u_{ij}^m} \quad (16)$$

Hesaplanan küme merkezlerine göre, U matrisi Eşitlik (17) ile yeniden hesaplanır. Eski U matrisi ile yeni U matrisi karşılaştırılır ve fark ε 'dan küçük olana kadar işlemler devam eder (Işık ve Çamurcu, 2007).

$$u_{ij} = \frac{1}{\sum_{k=1}^S \left(\frac{\|x_i - c_k\|}{\|x_i - c_j\|} \right)^{2/(m-1)}} \quad (17)$$

Kümeleme işlemi sonucunda bulanık değerler içeren U üyelik matrisi kümelemenin sonucunu yansıtır. İstenirse, berraklaştırma yapılarak bu değerler yuvarlanıp 0 ve 1'lere dönüştürülebilir (Işık ve Çamurcu, 2007).

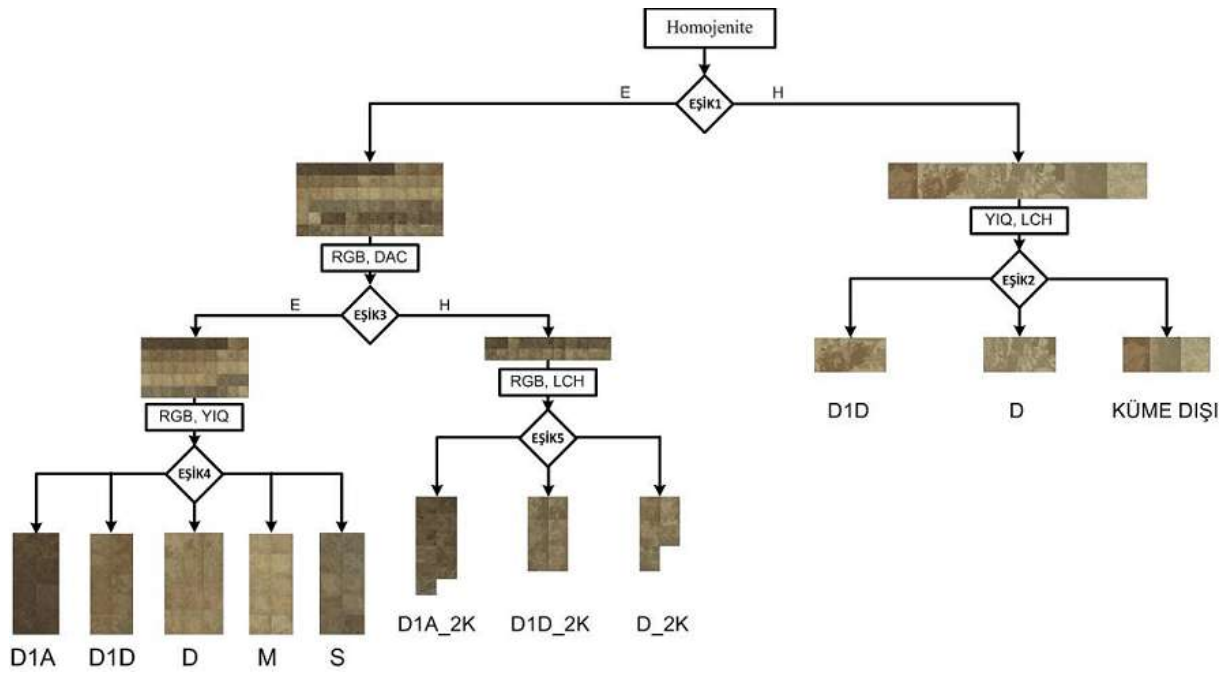
Bulgular

Deneysel çalışmada Başarırlar Mermer Sanayi'den elde edilen 45x45 cm boyutlarındaki 80 adet traverten örneklerinin, ARTEM Otomasyon tarafından hazırlanan otomatik mermer tanıma sistemi ile görüntüleri elde edilmiştir. Traverten veri tabanının üretici tarafından belirlenen beş adet birinci kalite ve üç adet ikinci kalite sınıfları bulunmaktadır. Bunlar sırasıyla D1A, D1D, D, M, S, D1A_2K, D1D_2K, D_2K sınıf etiketlerine sahiptirler. Her bir resim için toplam 227 adet öznitelik çıkarılmıştır. Bunlar RGB, CMYK, YIQ, YCbCr, IND, UPVPL, UVL, XYL, XYZ, YPBPR, YDBDR, YUV, HSV, HSL, HSI, DAC, CIE-L*a*b*, CIE-L*c*h, CAT02-LMS, OHTA, JPEG-YCbCr renk uzaylarının ortalama, standart sapma değerleri ve Gabor Filtre değerleridir (189 adet). Ayrıca RGB resimlerde her bir traverten üçer kümeye ayrılarak K-ortalamalar yöntemiyle elde edilen küme merkezleri, küme standart sapmaları ve kümelerin birbirlerine olan uzaklıkları elde edilerek sınıf tespitinde kullanılmıştır (18 adet). Ayrıca RGB renk uzayında zıtlık, HM, ilinti, enerji, entropi değerleri de elde edilmiştir (20 adet).

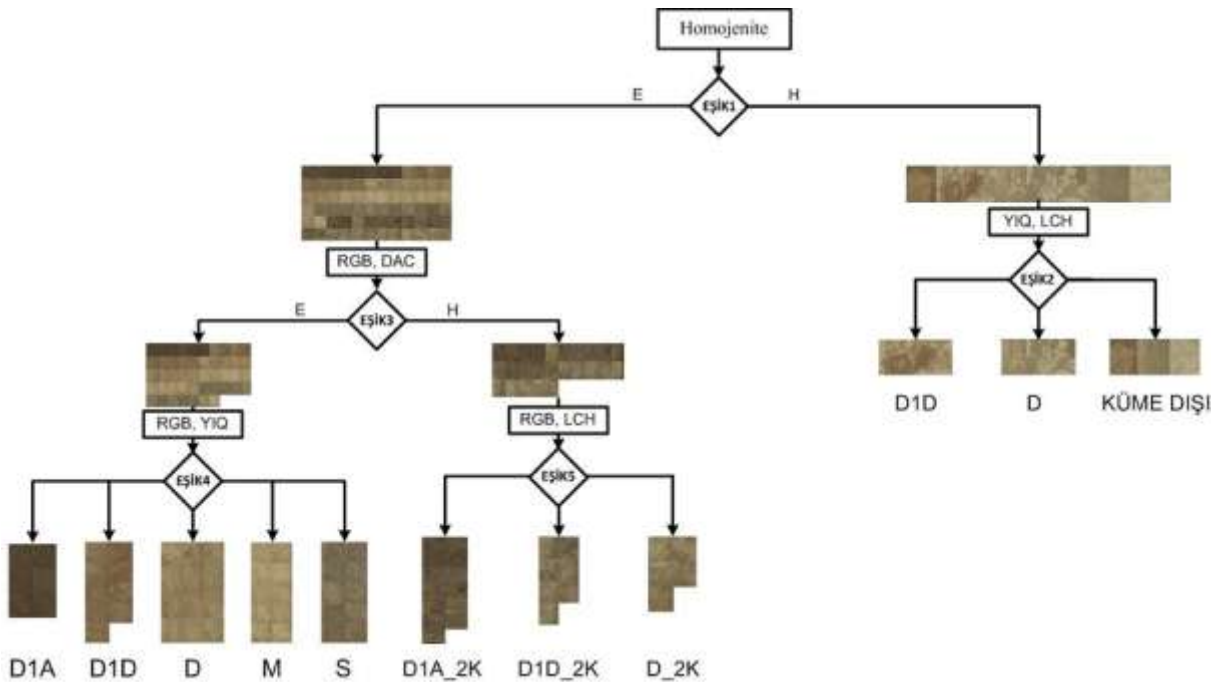
Elde edilen öznitelikler Pearson ilinti ölçütü ile ilk önce yediye indirilmiştir. Bunlar RGB renk ortalama değerleri, HM, YIQ renk uzayında I ortalama, CIE-L*c*h renk uzayında ise h , DAC renk uzayında ise I_d bileşenin standart sapmasıdır.

Bu çalışmada, seçilen öznitelikler birer birer K-ortalamalar ve bulanık C-ortalamalar yöntemleri ile değerlendirilmiş ve elde edilen sonuçlarla bir karar ağacı oluşturulmaya çalışılmıştır. K-ortalamalar ve bulanık C-ortalamalar için oluşturulan karar ağaçları sırasıyla Şekil 3 ve Şekil 4'te gösterilmektedir. Karar ağacının tepesinde renk değerlendirmesinden önce HM'ye bakılarak, zıtlığı fazla olan ikinci kalite travertenler, birinci kaliteden ayrıştırılmıştır. Şekillerde HM için verilen Eşik1'in değerleri aşağıdaki Tablo 1 ve Tablo 2'de verilmektedir.

Ardından oluşan her iki dalda da farklı öz niteliklerle yeni kümelemeler yapılmıştır.



Şekil 3. K-ortalamlar karar ağacı



Şekil 4. Bulanık C-ortalamlar karar ağacı

Ağacın sağ tarafında sadece ikinci kalite travertenler bulunup ilk önce YIQ renk uzayının I bileşeninin ortalama değeri ve CIE- L^*c^*h renk uzayının h bileşeni kullanılmıştır. Bu özelliklere ait sınırlar Eşik2 ile belirtilmektedir. Böylece herhangi bir sınıfa ait olmayan küme dışı örnekler belirlenmiştir. Bu örnekler düzgün ayrılan kümelerin dağılımlarını olumsuz etkilemektedirler.

Ağacın sol tarafında hem birinci hem de ikinci kalite travertenler bulunmaktadır. Bunların iki sınıfa ayrılması için ise RGB renk ortalama değerleri ve DAC renk uzayında I_d bileşeninin standart sapması kullanılarak kümeleme işlemi gerçekleştirilmiştir. Dallanmalarda kullanılan öz nitelik sınırları Eşik3 ile gösterilmektedir. Bu sayede ağacın sol tarafında iki dal oluşturulup soldaki dal birinci kalite travertenleri gösterirken sağ taraftaki dal ise ikinci kalite travertenleri göstermektedir. Birinci kalite

travertenlerin bir arada bulunduğu sol taraftaki dalda kendi içinde beş alt dala ayrılmaktadır. Ağacın yaprakları olan bu beş sınıf kümeleme ile belirlenmiştir; kümelemede RGB renk ortalama değerleri ve YIQ renk uzayının I bileşeninin ortalama değeri kullanılmıştır. Tanımlanan sınırlar Eşik4 ile belirtilmektedir. Ağacın yaprakları D1A, D1D, D, M ve S sınıf etiketlerine Bu kümeleme işlemi sonrasında bulunan küme eşik değerleri ya da aralıkları her iki kümeleme yöntemi için Tablo 1 ve Tablo 2’de belirtilmektedir.

sahiptir. İkinci kalite travertenlerin bulunduğu sağ taraftaki dal ise kendi içinde RGB renk ortalama değerleri ve CIE-L*a*b* renk uzayının h bileşeni kullanılarak üç yaprağa ayrılmıştır. Bu üç yaprak D1A_2K, D1D_2K, D_2K sınıf etiketlerine sahiptir. Tanımlanan sınırlar ise Eşik5 ile gösterilmektedir.

Tablo 1. K-ortalamalar eşik değerleri

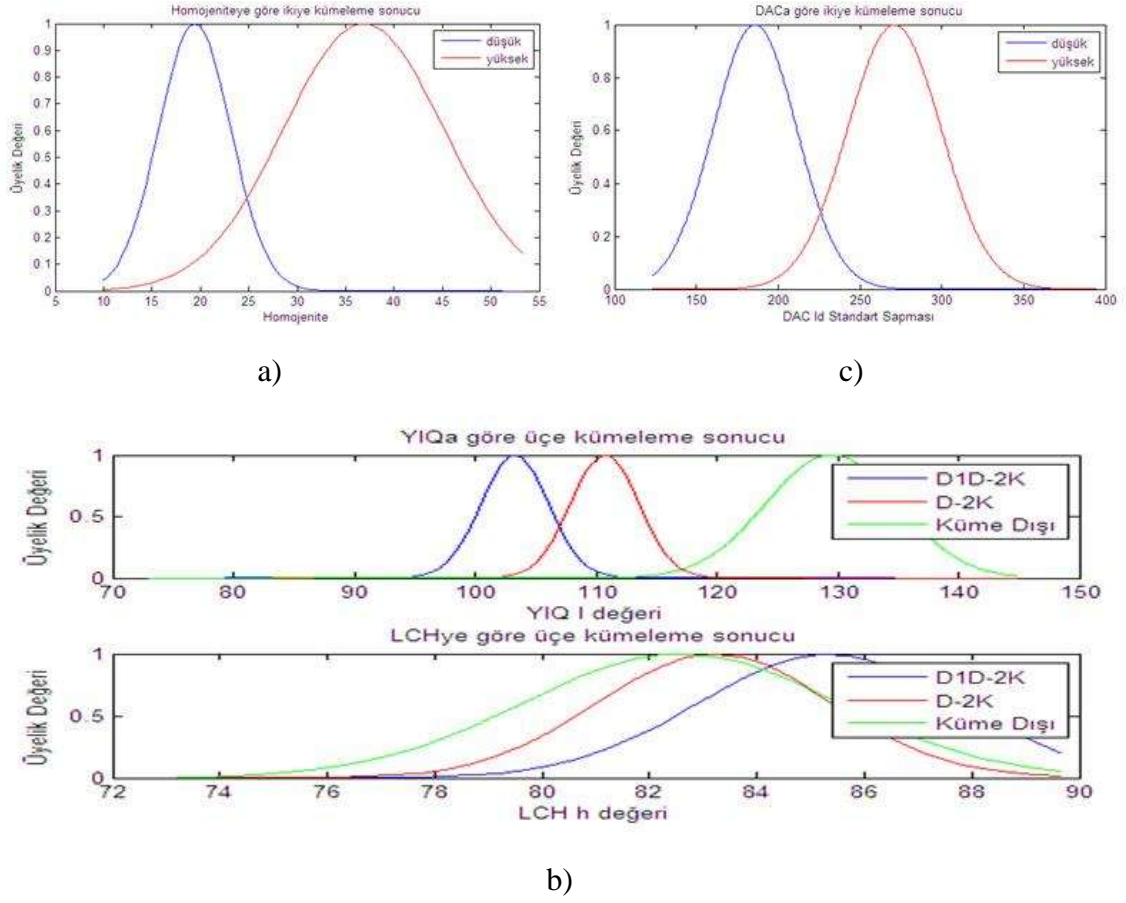
Öznitelikler Etiketler	R	G	B	HM	YIQ	LCH	DAC	Sınıfı
Eşik1	-	-	-	24.91	-	-	-	Homojenlik
Eşik2					117	83.2	-	Küme Dışı
Eşik3	-	-	-	-	-	-	230	1.-2. Kalite
Eşik4	$R \leq 100$	$G \leq 81$	$B \leq 59$	-	≤ 103	-	-	D1A_1K
	$100 < R \leq 106$	$87 < G \leq 92$	$60 < B \leq 74$		$72 < I \leq 108$			D1D_1K
	$105 < R \leq 135$	$103 < G \leq 120$	$72 < B \leq 89$		$87 < I \leq 112$			D_1K
	$135 < R \leq 170$	$119 < G \leq 160$	$89 < B \leq 120$		$120 < I \leq 150$			M_1K
	$114 < R \leq 133$	$90 < G \leq 103$	$70 < B \leq 79$		$116 < I \leq 118$			S_1K
Eşik5	$R \leq 101$	$G \leq 85$	$B \leq 63$				$259 < I_d \leq 298$	D1A_2K
	$100 < R \leq 115$	$85 < G \leq 100$	$59 < B \leq 76$				$300 \leq I_d$	D1D_2K
	$114 \leq R$	$101 \leq G$	$75 \leq B$				$I_d \leq 260$	D_2K

Tablo 2. Bulanık C-ortalamalar eşik değerleri

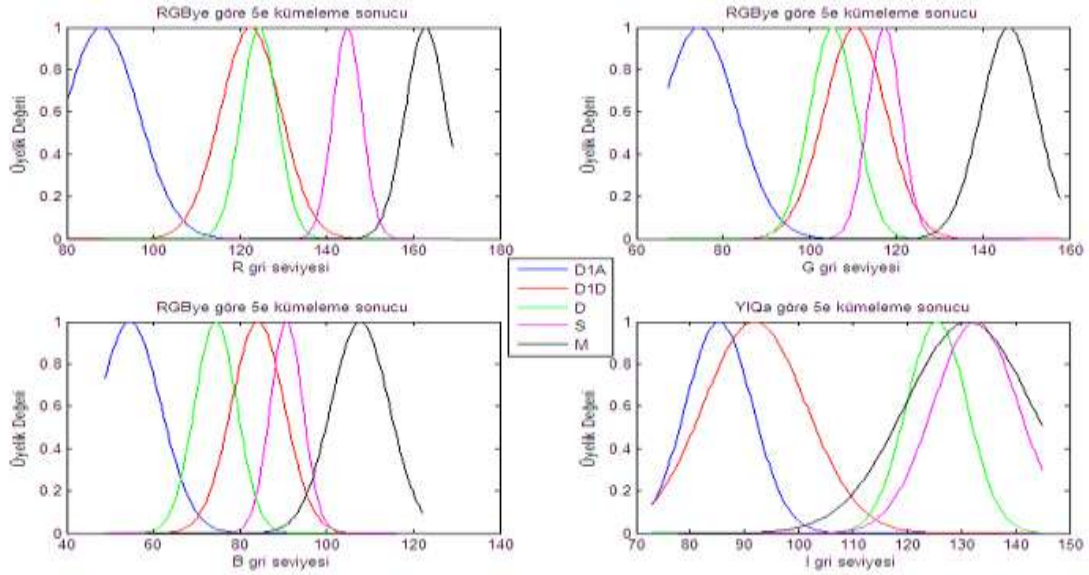
Öznitelikler Etiketler	R	G	B	HM	YIQ	LCH	DAC	Sınıfı
Eşik1	-	-	-	25	-	-	-	Homojenlik
Eşik2					117	83	-	Küme Dışı
Eşik3	-	-	-	-	-	-	223	1.-2. Kalite
Eşik4	$R \leq 94$	$G \leq 81$	$B \leq 60$	-	≤ 87	-	-	D1A_1K
	$95 < R \leq 123$	$80 < G \leq 118$	$78 < B \leq 88$		$86 < I \leq 107$			D1D_1K
	$123 < R \leq 127$	$101 < G \leq 107$	$60 < B \leq 79$		$107 < I \leq 127$			D_1K
	$152 < R$	$130 < G$	$97 < B$		$107 < I \leq 132$			M_1K
	$138 < R \leq 152$	$118 < G \leq 129$	$88 < B \leq 97$		$112 < I \leq 136$			S_1K
Eşik5	$R \leq 100$	$G \leq 85$	$B \leq 63$				$257 < I_d \leq 298$	D1A_2K
	$100 < R \leq 115$	$86 < G \leq 101$	$63 < B \leq 75$				$298 \leq I_d$	D1D_2K
	$114 \leq R$	$101 \leq G$	$75 \leq B$				$I_d \leq 257$	D_2K

K-ortalamalar eşik değerlerine göre sınıf ya da küme dağılımları aşağıdaki şekillerde gösterilmektedir. Şekil 5’de Eşik1, Eşik2 ve Eşik3’e göre dağılımlar gösterilmektedir. Şekil 6 ve Şekil 7’de ise Eşik4 ve

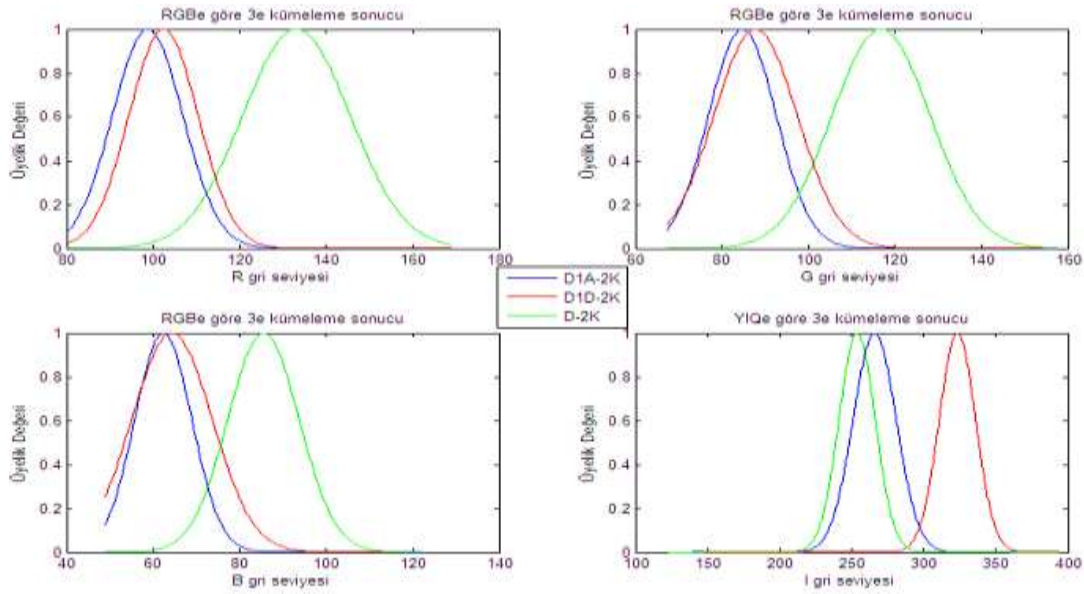
Eşik5’e ait dağılımlar sırasıyla verilmektedir. Elde edilen öznitelikler ve bu özniteliklerin kullanılma sırası üretici tarafından belirlenen sınıfların kendiliğinden ortaya çıkmasını, küme dışı örneklerin tespitini göstermektedir.



Şekil 5. Travertenlerin K-ortalamalara göre kümelenmesi a) Eşik1, b) Eşik2, c) Eşik3



Şekil 6. Travertenlerin Eşik4'e göre K-ortalamalar ile kümelenmesi



Şekil 7. Travertenlerin Eşik5'e göre K-ortalamlar ile kümelenmesi

Veri setindeki traverten plaka taşlar insan gözüyle sınıflara ayrılmış olup, sınıflar içinde uyumsuzluklar göze çarpmaktadır. Bu yüzden traverten plaka taşlar orijinalinde belirlenen sınıflardan daha farklı şekilde ele

alınmıştır. Tablo 3'de K-ortalamlar kümeleme sonucu elde edilen veriler ile insan gözüyle sınıflandırma sonuçları başarı oranları görülmektedir.

Tablo 3. Travertenlerin elde edilen sınıf karşılıkları

Yeni Sınıflar	D1A	D1D	D	M	S	D1A_2K	D1D_2K	D_2K	Zıtlığı Fazla	Küme Dışı
Eski Sınıflar										
D1A	%80					%20				
D1D		%90					%10			
D			%100							
M			%10	%90						
S					%90				%90	%10
D1A_2K					%10	%90				
D1D_2K		%10					%60		%20	%10
D_2K			%10	%10				%40	%30	%10

Tablo 3'de görüldüğü gibi genel olarak traverten plaka taşlar %80 doğrulukla eski sınıflarına yerleştirilmektedir. Birinci kalite sınıflarda yerleştirme oranı %90 iken ikinci kalitede bu oranın %63'lere kadar düştüğü görülmektedir.

Buna göre ikinci kalite tespitlerinde büyük hatalar yapılmaktadır. Yapılan bu çalışma ile örneklerin yanlış sınıflara ayrılması önlenmekte ve sınıf içi dağılımların homojen olması sağlanmaktadır.

Tartışma ve Sonuç

Bu çalışmada, traverten plaka taşlar için öğreticisiz öğrenme yöntemlerinden biri olan K-ortalamlar kullanılarak kişilerin gözlemlerinden bağımsız bir sınıf tanımlaması yapılmıştır. Çeşitli renk ve HM özelliklerini içeren traverten taşları çoğu zaman şirketler tarafından el yordamı ile ayrılmaya çalışıldığından standart bir ayrıştırma yapılamamaktadır. Bunun en önemli sebeplerinden biride hiç kuşkusuz kişiden kişiye değişen renk gözlemidir. Bu sebeple bu çalışmada bilindik RGB (Kırmızı-Yeşil-Mavi) renk uzayına ek olarak YIQ, CIE-L*c*h ve DAC renk uzayları ile HM ölçütü de birer öznitelik olarak kullanılmıştır. Deneysel çalışmalardan da

görüldüğü gibi K-ortalamlar kümeleme yönteminin yanı sıra kullanılan bulanık C-ortalamlar kümeleme yöntemi de yaklaşık aynı sonucu vermektedir. Elde edilen sonuçlar ışığında geliştirilen programın özniteliklerinin iyi seçildiği ve programda önemli bir yere sahip olup kümeleme yöntemlerinden neredeyse bağımsız olduğu görülmektedir. Bu çalışmada ayrıca hiçbir sınıfa ait olmayan traverten taşların da bulunması sağlanmıştır. Bu sayede gerek sınıfların belirlenmesinde gerek dekorasyon olarak yapılan çalışmalarda birkaç taş yüzünden görselliğin bozulması önlenmektedir. Bu az sayıdaki taşlar kişiler tarafından çok iyi belirlenemeyip çoğunlukla

herhangi bir sınıfa dâhil edilmektedir. Deneysel çalışmalarda da görüldüğü gibi daha kaliteli ve daha iyi sonuç veren bu program sayesinde insan hatalarının önüne geçilip çok daha hızlı ve standart sonuçlar veren bir sistem geliştirilmiştir.

Kaynaklar

- Benavente, N., Pina, P. 2009. Morphological segmentation and classification of marble textures at macroscopical scale, *Computers & Geosciences*, 35(6), 1194-1204.
- Castleman, K.R. 1996. *Digital Image Processing*, Prentice Hall, New Jersey.
- Chaves-González, J.M., Vega-Rodríguez, M.A., Gómez-Pulido, J.A., Sánchez-Pérez, J.M. 2010. Detecting skin in face recognition systems: A colour spaces study, *Digital Signal Processing*, 20(3), 806-823.
- Dunn, J.C. 1973. A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact, Well-Separated Clusters, *J. Cybern.*, 3, 32-57.
- Eraslan, H., İpçioğlu, İ., Haşit, G., Erşahan, B. 2008. Bilecik bölgesi mermer sektörünün uluslar arası rekabetçilik analizi: sektörel sorunlar ve çözüm önerileri. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 5(10), 193-217.
- Gonzalez, R.C., Woods, R.E. 2008. *Digital Image Processing*, Pearson Int. Ed., 2008.
- Han, J. ve Kamber, M. 2006. *Data Mining Concepts and Techniques*, Morgan Kauffmann Publishers Inc.
- Hopfield, J.J. 1982. Neural Networks and Physical Systems with Emergent Collective Computational Abilities, In *Proc. of National Academy of Sciences*, 79, 2554-2558.
- Höppler, F., Klawoun, F., Kruse, R., Ruukler, T. 1999. *Fuzzy Cluster Analysis*, Wiley.
- Işık, M., Çamurcu, A.Y. 2007. K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti, *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11, 31-45.
- Jang J-S.R., Sun C-T., Mizutani E. 1997. *Neuro-fuzzy and soft computing*, Prentice Hall, Upper Saddle River, 403-453 pp.
- Karaca, Z., Öztank, N. 2003. Mermerin mimari tasarımı, *Türkiye IV. Mermer Sempozyumu (MERSEM'2003)*, 18-19 Aralık, Afyonkarahisar, 379-386.
- Kohonen, T. 1977. *Associative memory: a system-theoretical approach*, Spring-Verlag, New York.
- López, M., Martínez, J., Matías, J.M., Taboada, J., Vilán, J.A. 2010. Functional classification of ornamental stone using machine learning techniques, *J. of Computational and Applied Mathematics*, 234(4), 1338-1345.
- MacQueen, J.B. 1967. Some Methods For Classification and Analysis of Multivariate Observations, *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 281-297.
- Martinez-Alajarin, J., Luis-Delgado, J.D., Tomas-Balibrea, L.M. 2005. Automatic system for quality-based classification of marble textures, *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 35(4), 488-497.
- Metlek, S., Büyükçingir, E., Cetişli, B. 2009. The automatic machine vision and classification system of granite slabs, *Int. Symposium on Eng. and Arc. Sciences of Balkan, Caucasus and Turckic Republics*, 22-24 Ekim, Isparta, 99-103.
- Osareh, A., Mirmehdi, M., Thomas, B., Markham, R. 2002. Comparison of Colour Spaces for Optic Disc Localisation in Retinal Images, *IEEE 16th Int. Conf. on Pattern Recognition*, 1, 743-746.
- Özdamar K. 2004. Paket Programlar İle İstatistiksel Veri Analizi 2, Kaan Kitabevi, 5. Baskı, Eskişehir, Türkiye, 213-230 ve 279-292.
- Özkan, Y. 2008. *Veri Madenciliği Yöntemleri*, Papatya Yayıncılık ve Eğitim AŞ. Bölüm 6.
- Periavi, A., Kheibari H.T. 2008. A Fast Algorithm For Connectivity Graph Approximation Using Modified Manhattan Distance in Dynamic Networks, *Applied Mathematics and Computation*, 201, 319-332.
- Selver, A.M., Akay, O. 2009. Evaluating clustering methods for classification of marble slabs in an automated industrial marble inspection system, *Int. Conf. on Electrical and Electronics Engineering, ELECO 2009*, II-115 - II-119.
- Taşkın, Ç., Emel, G.G. 2010. Veri Madenciliğinde Kümeleme Yaklaşımları Ve Kohonen Ağları İle Perakendecilik Sektöründe Bir Uygulama, *Süleyman Demirel Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi*, 15, 395-409.
- Theodoridis, S., Kotroumbas, K. 1999. *Pattern Recognition*. Academic Press, San Diego.
- Viénot, F., Brettel, H. & Mollon, J. D. 1999. Digital Video Colourmaps for Checking the Legibility of Displays by Dichromats, *Color Research and Application*, 24, 243-252.