Sıradan Metinlerin İçinde **Gizli Bilgiler**

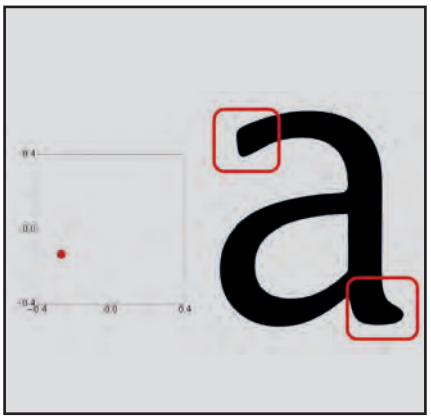
Dr. Mahir E. Ocak [*TÜBİTAK Bilim ve Teknik Dergisi*

Columbia Üniversitesi'nde çalışan bir grup araştırmacı, sıradan metinlerin içine bilgiler gizlemek için yeni bir yöntem geliştirdi. FontCode adı verilen teknik, yazı tipinde çıplak gözle fark edilemeyecek değişiklikler yapılmasına dayanıyor. Dosyalara gizlenen bilgiler, belge yazdırıldığında ya da başka formatlara dönüştürüldüğünde kaybolmuyor. Görünümlerini ya da düzenlerini değiştirmeden belgelere QR kodları ya da başka bilgiler eklemek, telif haklarını korumak ya da belgelerin içeriklerinin değiştirilmesini önlemek amacıyla bu tekniğin kullanılması mümkün.

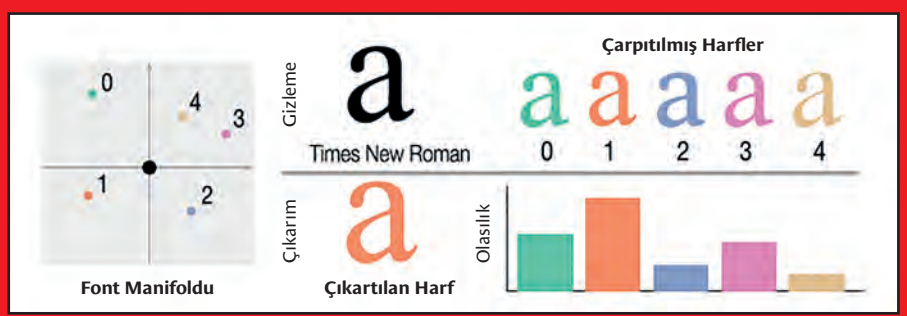
Verilerin içine bilgi gömme tekniklerinin pek çok uygulaması var. Örneğin dijital fotoğraf dosyalarının içinde, fotoğrafı çeken kameralar tarafından otomatik olarak yerleştirilmiş, kameranın GPS konumuyla, pozlandırma süresiyle ve odak mesafesiyle ilgili gömülü bilgiler bulunur. Resim, video ve ses dosyalarının üzerindeki telif haklarını korumanın en etkin yollarından biri de dosyalara filigran olarak adlandırılan gizli damgalar yerleştirmektir.

Tüm uygulama alanlarında bilgi gömme tekniklerinin sağlaması arzu edilen iki kriter olduğu söylenebilir. Birincisi gömülü mesajın dosyadaki verileri mümkün olduğu kadar az çarpıtması (değiştirmesi, bozması). İkincisi gömülü mesajın gerekli olduğunda okunabilmesi. Bu kriterleri sağlamanın en zor olduğu alan, metin dosyalarıdır. Ses ve görüntü dosyalarında kolay fark edilemeyecek ufak değişiklikler yapmak mümkündür. Örneğin seslerin frekanslarında ve renklerde insanların aradaki farkı algılayamayacağı kadar ufak değişiklikler yapılabilir. Ancak metin dosyalarının "pikselleri" harflerdir. Bir harf, farkına varılmayacak biçimde başka bir harfle değiştirilemez. Günümüzde metin dosyalarına bilgi gömmek için kullanılan çeşitli yöntemler olsa da hem kapasiteleri sınırlı hem de sadece belirli dosya formatlarında kullanılabiliyorlar.

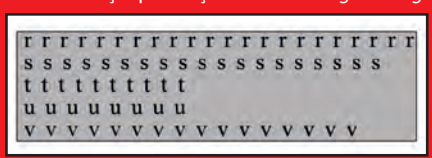
Örneğin PDF ya da Word dosyalarındaki boşlukların büyüklüklerini değiştirerek "0"lar ve "1"ler kodlanabiliyor ve bir metnin içine mesaj yerleştirilebiliyor. Ancak dosya başka formatlara çevrildiğinde içindeki mesaj yok oluyor.

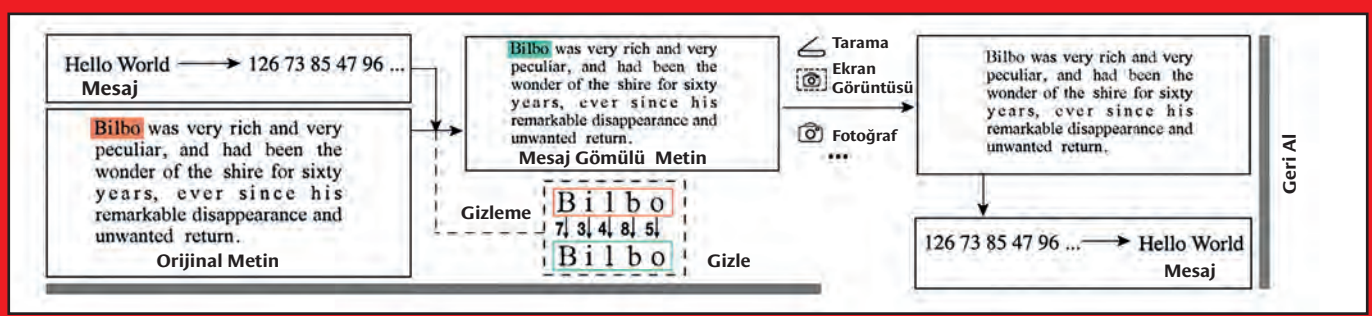


Columbia Üniversitesi araştırmacılarının geliştirdiği yeni bilgi gömme yönteminde yazı karakterlerinde çıplak gözle farkına varılamayacak ufak değişiklikler yapılıyor. Böylece metnin okunurluğunda ve görünümünde önemli bir değişiklik olmuyor. Dosyanın içine gömülen mesaj, çizgi kalınlığı değiştirilerek; harflerin alt ve üst çıkıntıları uzatılarak ya da kısaltılarak; o, p, b gibi harflerdeki yuvarlaklıklar kalınlaştırılarak ya da inceltilerek kodlanıyor. FontCode adı verilen yöntemi Times New Roman, Helvetica, Calibri gibi yaygın kullanılan fontlarla; Work, FrameMaker gibi metin işleme programlarıyla ve Photoshop, Illustrator gibi resim işleme ve çizme programlarıyla kullanmak mümkün. Ayrıca dosya başka formatlara dönüştürüldüğünde içindeki mesaj yok olmuyor. Metinlerdeki her harfin biçiminde ufak değişiklikler yapılabildiği için dosyaların içine gömülecek mesajın uzunluğunu sınırlandıran tek şey metnin uzunluğu. Her ne kadar fontlarda yapılan değişiklikler çıplak gözle fark edilemese de makineler yardımıyla gömülü mesajların okunması mümkün.



Çarpıtılmış harfler kataloğu örneği





FontCode'un **İşleyiş Biçimi**

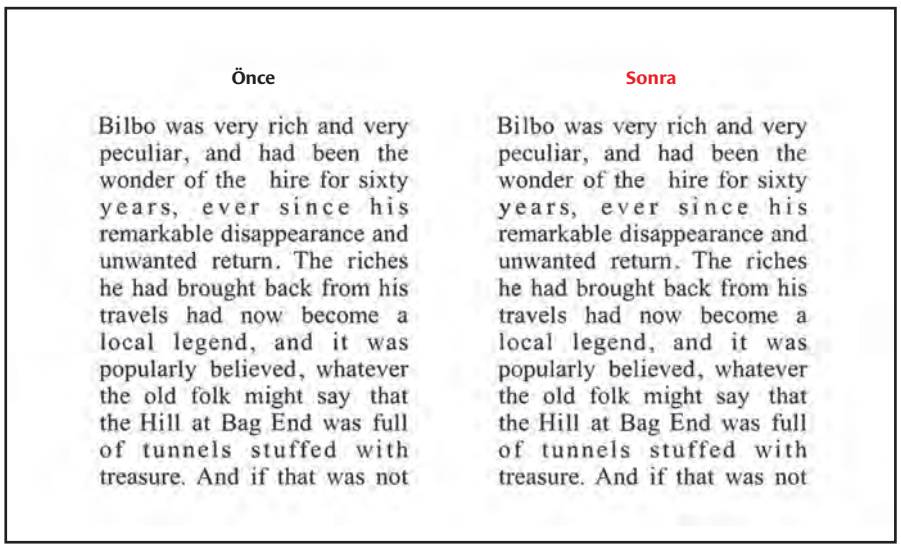
Mesaj yerleştirme

Metinlerin içine mesaj gömme işlemi, çarpıtılmış fontlar içeren bir katalog hazırlanması ve mesajların metne yerleştirilmesi şeklinde iki aşamada ele alınabilir.

Çarpıtılmış fontlar kataloğu hazırlanırken "font manifoldu" düşüncesinden yararlanılıyor (*bkz. bir önceki sayfada yer alan şekil*). Örneğin Times New Roman ya da Helvetica gibi yaygın olarak kullanılan bir fontu ele alalım. Bu fontlardaki her bir yazı karakteri (harfler, noktalama işaretleri vs.), font manifoldunun belirli bir noktasında bulunuyor. Bir yazı karakterinin çarpıtılmış halleri hazırlanırken önce o karakterin font manifoldunda bulunduğu noktanın etrafında birkaç nokta seçiliyor. Yeni noktaların karşılık geldiği çarpıtılmış karakterlerin orijinal karakterden farkının çıplak gözle ayırt edilemeyecek kadar küçük ancak makineler yardımıyla ayırt edilebilecek kadar büyük olanları seçiliyor. Daha sonra bu çarpıtılmış karakterlere farklı sayılar atanarak fontlar kataloğa ekleniyor.

Metnin içine mesaj yerleştirilirken ilk olarak dijital bilgisayarlarda olduğu gibi mesaj sayılarla ifade ediliyor. Daha sonra metindeki bir harfe bir sayıyı kodlamak için o harfin çarpıtılmış fontlar kataloğunda o sayıya karşılık gelen formu kullanılıyor. Örneğin a harfinin beş ayrı çarpıtılmış formu olduğunu düşünelim. Bu formların her biri 0, 1, 2, 3, 4 sayılarından birine karşılık geliyor. Eğer a harfine 2 sayısı kodlanmak isteniyorsa, orijinal metindeki hatasız a harfi 2 sayısına karşılık gelen çarpıtılmış a harfiyle değiştiriliyor.

Metinlerin içindeki gömülü mesajların ortaya çıkarılması sürecindeyse işlemlerin sırası tersine dönüyor. İlk olarak metindeki çarpıtılmış harfler ve bu harflerin karşılık geldiği sayılar okunuyor. Daha sonra dijital olarak gizlenmiş mesaj yazılı hale çevriliyor. Mesajların hatalı okunması da mümkün, ancak araştırmacılar hata düzeltme algoritması da geliştirmişler.



Mesaj gömme ve okuma süreçlerinin düzgün bir biçimde işlemesi için gerekli en önemli şey, çarpıtılmış harflerin tanımlanması ve birbirinden ayırt edilmesi. PDF dosyaları ve Illustrator gibi programlarla hazırlanmış dosyalar için bu işlem hayli kolaydır. Çünkü bu dosyalardaki harfler ve şekiller kolayca bilgisayarlar tarafından okunabilir. Ancak PNG, IMG ve diğer pikselleştirilmiş dosyalardaki çarpıtılmış harfleri ayırt etmekse çok daha zordur. Çünkü ışık miktarındaki ya da kamera açısındaki ufak farklar veya görüntüdeki bir miktar bulanıklık harflerin ayırt edilmesini zorlaştırır. Araştırmacılar, bu zorluğu aşmak ve geliştirdikleri yöntemin herhangi bir dosya formatıyla kullanılabilmesini sağlamak için kısaca CNN olarak adlandırılan bir yapay zekâ türüne yönelmişler. Programın çarpıtılmış harfleri tanıması için, önce bir metnin içindeki çarpıtılmış harfler tek tek kesilip çıkarılıyor ve bu görüntüler dijital hale getiriliyor. Daha sonra görüntü 200x200 piksel boyutlarında siyah-beyaz bir resim haline getirilerek CNN'leri eğitmek için kullanılıyor. Araştırmacılar, eğitim aşamasında her bir çarpıtılmış harfi kâğıda yazdırıp farklı ışık koşullarında farklı açılardan 10'ar fotoğraf çekmişler. Ayrıca bilgisayar kullanılarak üretilen sentetik görüntüler de eğitim setine eklenmiş. Eğitimler sırasında CNN'lere bu setten rastgele seçilmiş görüntüler gösteriliyor. Dolayısıyla CNN'ler sadece gizlenmiş metinlerin okunmasında değil, gizli metinleri yerleştirirken kullanılacak fontları içeren kataloğun oluşturulmasında da rol alıyorlar.

FontCode'un **Uygulamaları**

Formattan bağımsız üst veri

Pek çok dijital ürün üst veri içerir. Örneğin resim dosyalarının içinde resmi çeken kamerayla ilgili üst veriler bulunur. Bugün dosyalara üst veri eklemek için kullanılan yöntemlerle ilgili en önemli sorun, dosya formatına bağlı olmaları. Dolayısıyla dosya formatında bir değişiklik yapıldığında üst veriler kayboluyor. Örneğin JPEG dosyası PNG dosyasına çevrildiğinde ya da vektör grafik dosyaları piksel grafik dosyalarına çevrildiğinde orijinal dosyalardaki üst veriler kayboluyor. Bugün kullanılan yöntemlerle karşılaştırıldığında FontCode'un en önemli avantajı dosya formatından bağımsız olması. Bir kez üst veriler eklendikten sonra dosyayı güvenle başka formatlara dönüştürmek, yazdırmak ya da pikselleştirmek mümkün. Çünkü bu işlemler sırasında çarpıtılmış yazı karakterlerinin şekli değişmediği için üst veriler korunuyor.

Algılanamayan optik kodlar

Ticaret, reklamcılık, artırılmış gerçeklik gibi pek çok alanda optik barkodlar kullanılıyor. FontCode'u da yazı dosyalarının içine barkodlar eklemek için kullanmak mümkün. Günümüzdeki barkodların tamamı siyah-beyaz çubukların ya da blokların yazdırılmasına dayanıyor. Dolayısıyla barkodların görünümü rahatsız edici olabiliyor. FontCode ise yazı karakterlerinde çıplak gözle fark edilemeyecek değişiklikler yapılmasına dayalı bir yöntem olduğu için görünümde önemli bir değişikliğe sebep olmuyor. Özellikle poster ya da broşür tasarımı gibi görselliğin önemli olduğu alanlarda bugün kullanılan QR kodlarının yerini FontCode alabilir. Araştırmacılar, metinlerin içinde gizlenmiş QR kodlarını okumak için bir akıllı telefon uygulaması da geliştirmişler.

Şifreli mesajlar

FontCode ile yazıların içine yerleştirilen mesajların şifreli hale getirilmesi de mümkün. Birbirine şifreli mesaj göndermek isteyen iki kişi, içeriğini herkesin bildiği, ortak bir çarpıtılmış font kataloğu kullanabilir. Hangi çarpıtılmış harfin hangi sayıya karşılık geldiğini ise kendi aralarında belirleyip bir sır olarak tutabilirler. Böylece, her ne kadar herhangi biri hangi harflerin çarpıtılmış olduğunu makineler yardımıyla ayırt edebilse de hangi harfin hangi sayıya karşılık geldiğini bilemeyeceği için şifreli mesajların okunması imkânsız hale gelir.

Metinlerin değiştirilmesini engellemek

FontCode, metinlerin içeriğinin değiştirilmesini engellemek için de kullanılabilir. Örneğin bir yazar kriptografik yöntemler kullanarak, kaleme aldığı yazının metninden bir bit dizisi üretebilir. Daha sonra bu bit dizisini hangi çarpıtılmış harfin hangi sayıya karşılık geldiğini kendisi belirleyerek metnin içine gömebilir. Dosyanın içeriğinde yapılacak herhangi bir değişiklik farklı bir bit dizisinin üretilmesiyle sonuçlanacağı ve hangi yazı karakterinin hangi sayıya karşılık geldiğini sadece orijinal metnin yazarı bildiği için iz bırakmadan metinde değişiklikler yapmak imkânsızlaşır. Yazar kendi metninde herhangi bir değişiklik yapılıp yapılmadığını kontrol etmek istediğinde elindeki dosya üzerinde kodlanmış sayıları bir makine yardımıyla okur ve orijinal dosyanınkiyle aynı olup olmadığını kontrol eder.

FontCode, metinlerin içine görünümlerini bozmadan bilgiler gömmeye imkân veriyor. Üstelik dosya başka formatlara çevrildiğinde ya da yazdırıldığında bilgiler kaybolmuyor. Gelecekte poster, broşür ve dergi yazılarının üzerine gözümüzle algılayamadığımız gizli bilgiler ve barkodlar yerleştirmek, metinlerin orijinalliğini kontrol etmek ya da kriptolojik amaçlar için FontCode'un kullanılması mümkün. Şu an için geliştirilen yöntem sadece belirli fontlarla çalışıyor. Ancak herhangi bir fontla kullanılacak şekilde geliştirilmesinin önünde hiçbir zorluk yok. Ayrıca şekilleri çarpıtarak mesaj yerleştirme düşüncesinin herhangi bir sembolik sisteme uygulanması da mümkün. Örneğin müzik notalarında ya da matematiksel sembollerde çıplak gözle farkına varılamayacak ufak değişiklikler yaparak da bilgi gömmek mümkün olabilir. Ayrıca FontCode, Çince gibi, metinlerin harflerle değil logogramlarla yazıldığı dillerle kullanılacak biçimde de geliştirilebilir. Ancak logografik dillerdeki yazı karakteri sayısının alfabetik dillerdekinden çok daha fazla olması pratik uygulamaları zorlaştırabilir. ■

Kaynak

Xiao, C. ve ark., "Font Code: Embedding Information in Text Documents Using Glyph Perturbation", *ACM Transactions on Graphics*, Cilt 37, No: 2, Makale no: 15, 2018.