

Zamanda Ortalaması Alınmış İkili Önplan İmgeleri Kullanarak Taşıt Sınıflandırması

Classification of Vehicles Using Binary Foreground Images Averaged Over Time

Hakkı Can Karaimer, Yalın Baştanlar
Bilgisayar Mühendisliği Bölümü, İzmir Yüksek Teknoloji Enstitüsü, İzmir, Türkiye
{cankaraimer, yalinbastanlar}@iyte.edu.tr

Özetçe—Bu bildiriye tümyönlü videolarda taşıt sınıflandırması için şekil tabanlı bir yaklaşım sunulmaktadır. Önerilen yöntem, arkaplan çıkarımı sonrası elde edilen hareketli nesnenin silüetini kullanmaktadır. Diğer yöntemlerden farklı olarak, ardışık video karelerinden elde edilen silüetlerin zamansal ortalaması kullanılmaktadır. Deneyler, ortalama silüet yönteminin, tek kareden çıkarılan silüet kullanmaya göre daha başarılı sınıflandırma yaptığını göstermektedir. Ayırt edilen araç tipleri; motosiklet, binek araç ve dolmuştur. Tümyönlü videolardan bir veri seti oluşturulmuş ve farklı eğitim ve test kümelerine bölünüp deney tekrarlanarak rastsallaştırma sağlanmıştır.

Anahtar Kelimeler — Tümyönlü kamera, Tümyönlü video, Taşıt tespiti, Taşıt sınıflandırması

Abstract—We describe a shape-based method for classification of vehicles from omnidirectional videos. Different from similar approaches, the binary images of vehicles obtained by background subtraction in a sequence of frames are averaged over time. We show with experiments that using the average shape of the object results in a more accurate classification than using a single frame. The vehicle types we classify are motorcycle, car and van. We created an omnidirectional video dataset and repeated experiments with shuffled train-test sets to ensure randomization.

Keywords — Omnidirectional camera, Omnidirectional video, Vehicle detection, Vehicle classification

I. GİRİŞ

Tümyönlü kameralar tek bir imgede, yatay ekseninde 360° görüş açısı sağlarlar. Standart bir kameranın önüne konveks bir ayna yerleştirilerek katadioptrik tümyönlü kamera oluşturulabilir ve Şekil 1’deki gibi görüntüler elde edilebilir.

Standart kamera görüntülerinde nesne tespiti için önerilen yöntemlerin bir kısmı gitgide büyüyen ve kayan bir pencere ile arama yapmaktadır. Gradyan yönü, gradyan büyüklüğü ve renk gibi öznitelikler sınıflandırma için kullanılır.

Bu bildiriye çalışmaları TÜBİTAK tarafından 113E107 nolu proje kapsamında desteklenmiştir.

Önemli bir performans iyileşmesi, bu yaklaşımın HOG (Histogram of Oriented Gradients – Yönlü Gradyan Histogramı) öznitelikleriyle kullanılması ile elde edilmiştir [1]. Yakın zamanlı bazı çalışmalarda kayan pencereler yöntemi tümyönlü kameralar üzerinde de uygulanmıştır. Örneğin, [2]’de HOG hesaplamaları tümyönlü kameralara uyarlanmıştır. Ayrıca Haar-benzeri öznitelikler de tümyönlü kameralar ile kullanılmıştır [2][4].

Nesne tespiti için diğer bir yaklaşım ise videodan arkaplan çıkarımı sonrasında elde edilen ikili imgeler üzerinde şekil tabanlı öznitelikler kullanmaktır. Örneğin [5]’te alan, genişlik, doluluk, uzanım, çevre, dışbükey örtü çevresi, uzunluk, bindirilen elipsin eksenleri ve momentlerden oluşan bir öznitelik kümesi kullanılmış, ardından doğrusal ayırtaç çözümlenmesi ile boyut indirilmesi yapılmıştır. Sonra ise nesnelere ağırlıklı k-en yakın komşu sınıflandırıcısı ile sınıflandırılmıştır. [6]’de konum, uzunluk, genişlik ve hız öznitelikleri ile nesnenin binek araç olması veya olmaması şeklinde bir sınıflandırma yapılmış, [7]’de konum, hız, elips eksenleri bir bayesçi ağda kullanılmıştır. [8]’de ise araçların 3B modellerinin imge üzerine izdüşümü sonucu örtüşmesi hesaplanmıştır. [9]’de video üzerinden sanal bir doğru kullanılarak çıkartılan “zaman-uzamsal imge” ve k-en yakın komşu algoritması kullanılmıştır. Tümyönlü kamera ve şekil tabanlı yöntem kullanan bulabildiğimiz tek çalışmada [10] ise sadece alan öznitelik olarak kullanılıp araçlar küçük ve büyük olarak ikiye ayrılmıştır.

HOG ya da Haar-benzeri öznitelikler kullanan imge tabanlı yöntemler ile ikili imgelerden çıkan öznitelikleri kullanan şekil tabanlı yöntemler karşılaştırıldığında, imge tabanlı yöntemlerde öznitelik çıkarırken kayan pencereler yöntemi kullandığından hafıza ve zaman karmaşıklığı fazladır. Bu yükü azaltmak için sadece hareketin olduğu bölgeye odaklanılsa bile yine de tek bir pencere yerleştirmek mümkün değildir. Örneğin, [11]’deki HOG tabanlı yöntemde öznitelik çıkarılacak pencereler elle işaretlenmiştir. Sözü edilen iki yaklaşımın performans karşılaştırması ise [11] ve [5] ile yapılabilir. [11]’deki HOG tabanlı yöntemin sınıflandırma başarısı [5]’teki şekil tabanlı yöntemin çok altında çıkmıştır.

Bu nedenlerden dolayı tümyönlü kameralar için şekil tabanlı bir yöntem geliştirmeye karar verdik. Çalışmamızın temel katkısı çok sayıda video karesinden gelen bilginin kullanılmasıdır. Ardışık karelerden elde edilen hareketli nesneye ait ikili imgeler, diğer adıyla silüetler, “ortalama silüet” oluşturmak amacıyla birleştirilir. Bu işlemin benzeri görüntü işlemede gürültü eleme amacıyla “ortalama imge” elde etmek için kullanılmaktadır. Diğer bir katkımız ise silüet alanının öznitelik olarak sınıflandırmada kullanılmamış olmasıdır. Bu şekilde, yöntemimiz taşınabilir görüntüleme platformları için uygun hale gelmiştir. Literatürde kameraların binalara sabitlendiği çalışmalarda ([5][8][9][10]), silüet alanı öznitelik olarak kullanılmıştır ki bu öznitelik kamera-nesne mesafesi değiştiğinde kullanılmaz hale gelir.

Üzerinde çalıştığımız taşıt tipleri motosiklet, binek araç ve dolmuştur. Silüetlerden elde ettiğimiz öznitelikler ise dışbükeylik, uzanım, dikdörtgensellik ve Hu momentleridir. Dışbükeylik bozulmuş silüetleri elemek için, uzanım motosikletleri diğer araçlardan ayırmak için, diğer iki öznitelik de binek araç - dolmuş ayrımı için kullanılmıştır. Önerilen yöntemin performansı, tek bir video karesinden elde edilen silüet kullanan yöntemin sonuçları ile karşılaştırılmıştır. Ortalama silüet kullanıldığında performans motosiklet sınıfı için %80’den %95’e, binek araç sınıfı için %78’den %98’e, dolmuş sınıfı için %81’den %83’e çıkmıştır.

Tümyönlü video veri setimiz, arkaplan çıkarımı sonrası elde edilen ikili imgeler ile beraber, cvrg.iyte.edu.tr adresinden indirilebilir. Bölüm 2’de silüet ortalama işlemi, Bölüm 3’te ise tespit ve sınıflandırma basamakları anlatılmıştır. Deneyler Bölüm 4’te, vargılar Bölüm 5’te verilmiştir.

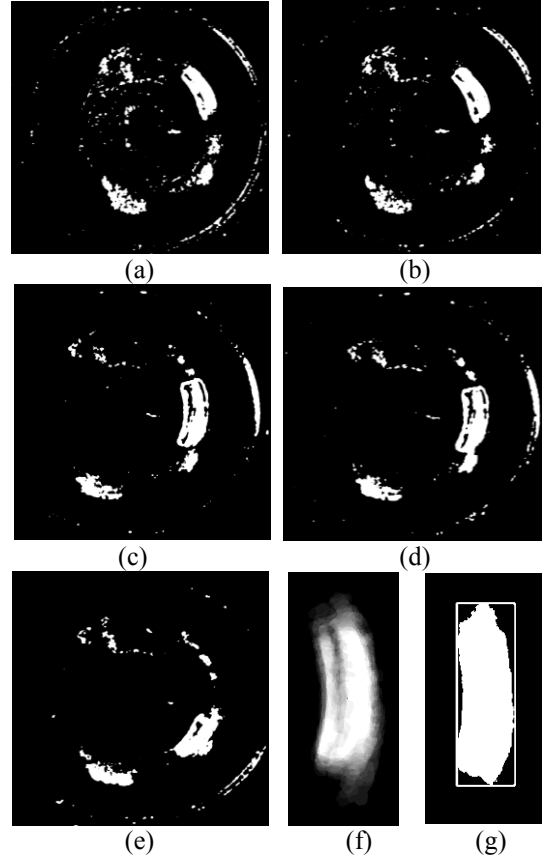
II. ZAMANDA ORTALAMASI ALINAN SİLÜETLER

Arkaplan çıkarımı ve morfolojik işlemler sonrası ikili önplan imgeleri elde edilir. Arkaplan çıkarım algoritması olarak [12]’de karşılaştırılan algoritmalarından, performansı en yüksek olanlardan biri olan [13] kullanılmıştır. Son maske ise bir disk ile yapılan açma işlemiyle elde edilir ve maske üzerindeki en büyük silüet hareket eden nesneye atanır. “Ortalama silüet” elde etmek için kaç video karesi kullanıldığı tanımlanmalı ve bu karelerden çıkan silüetler örtüştürülmelidir. Eğer bir silüet, önceden belirlenmiş açı aralığında (açı aralığı 0° yola en yakın nokta olacak şekilde $[30^\circ, -30^\circ]$ olarak seçilmiştir) bulunuyorsa o silüet tümyönlü imge merkezine göre döndürülür (Şekil 1). Silüet çıkarma işlemi, nesne açı aralığından ayrılıncaya kadar devam eder.

Önceki aşamada elde edilen silüetler, birbiri üzerine ağırlık merkezleri denk gelecek şekilde eklenir ve birikmiş imge oluşur. Birikmiş imgenin kullanılan kare sayısına bölünmesiyle ortalama silüet bulunmuş olur (Şekil 2f). Belirli bir eşik değerinin altındaki gri seviyelerinin elenmesiyle az sayıda video karesinin katkı verdiği bölgeler elenir ve sadece çok kare tarafından desteklenen parçalar dikkate alınmış olur (Şekil 2g). Burada seçtiğimiz eşik değeri, gri seviyelerinin aşağıdan %25’ini elemektedir.



Şekil 1: Üst: Yoldan geçen bir dolmuş içeren örnek tümyönlü video karesi. Alt-sol: Aynı karenin arkaplan ayırımından sonraki hali. Ayrıca imge üzerine bindirilmiş, kullandığımız açı aralığı, $[30^\circ, -30^\circ]$. Merkezi 29° ’de olan en büyük silüet. Alt-sağ: Morfolojik işlemlerden sonra döndürülmüş silüet.



Şekil 2: Örnek ikili imgelerde nesnenin merkezi (a) 29° ’de (b) 26° ’de (c) 0° ’de (d) -11° ’de (e) -29° ’de iken. (f) İkili imgelerdeki en büyük silüetlerden elde edilen ‘ortalama silüet’. (g) Eşik değeriyle işlenmiş silüet.

III. NESNE TESPİT VE SINIFLANDIRMASI

Şekil 4'teki akış şeması tespit ve sınıflandırma adımlarını özetlemektedir. Morfolojik işlemlere uğramış silüete ilk olarak dışbükeylik eşik değeri uygulanır. Burada, tek video karesi metodu için nesnenin 0° 'ye en yakın olduğu konumdaki silüet, ortalama silüet yöntemi için de Bölüm 2'deki işlemler sonucu oluşan silüet kullanılır.

Dışbükeylik, araç olmayan silüetleri ve araçlardan çıkan başarısız silüetleri elemek için önemli bir özneliktir [14].

$$Dışbükeylik = O_{Dışbükey örtü} / O \quad (1)$$

Burada $O_{Dışbükey örtü}$ dışbükey örtünün çevresi, O da orijinal konturun çevresidir. Tespit edilen silüet kümesinden $\{T_s\}$, geçerli olan tespitler $\{T_g\}$ bir dışbükeylik eşik değeri ile çekilebilir.

$$\{T_g\} = \{T_s | Dışbükeylik_{T_s} > \rho\} \quad (2)$$

Deneylerimizde ρ 'yu 0.75 olarak seçtik. Dışbükeylik eşik değeri ile elenmiş bir silüet Şekil 3'te görülmektedir. Geçerli tespitler ise sınıflandırma aşamasına geçerler.



Şekil 3. Bir dolmuş örneğine ait silüet. Dışbükeylik 0.73'tür. Bu da eşik değeri olan 0.75'in altındadır.

Sınıflandırmada kullanılan öznelilikler: uzanım, dikdörtgensellik ve Hu momentleridir. Uzanım (3) aşağıdaki gibi hesaplanır:

$$Uzanım = 1 - W/L \quad (3)$$

Burada W ve L sırasıyla silüeti çevreleyen en küçük dikdörtgenin kısa ve uzun kenarlarıdır [14]. Çevreleyen dikdörtgen örneği Şekil 2g'de verilmiştir.

Uzanımın, motosikletleri diğer araçlardan bir eşik değeri ile ayırabildiğini gözlemledik. Tespit edilen motosikletler kümesi $T_m(4)$ aşağıdaki gibidir:

$$\{T_m\} = \{T_m | Uzanım_{T_g} < \tau\} \quad (4)$$

Burada τ uzanım eşik değeridir. τ eğitim setindeki örnekler göre belirlenir.

Dikdörtgensellik (5), şeklin en küçük çevreleyen dikdörtgeni ne kadar doldurabildiğidir [14]:

$$Dikdörtgensellik = A_s / A_L \quad (5)$$

Burada A_s şeklin alanı, A_L ise çevreleyen dikdörtgenin alanıdır. Dikdörtgensellik, çevreleyen en küçük

dikdörtgeni doldurmaya daha yatkın olan dolmuşları ayırt etmek için uygun bir özneliktir. Denemelerimizde sadece dikdörtgensellik eşik değeri ile binek araç ve dolmuş ayrımı yapamadığımız için P_1 (8) adında Hu momentlerine dayanan bir öznelik tanımladık. Örneklere bağlı olan P_1 şu şekilde hesaplanır:

$$C_1 = \frac{1}{\#Binek araçlar} \sum_{i=0}^{\#Binek araçlar} I_2(T_g, Binek araç_i) \quad (6)$$

$$V_1 = \frac{1}{\#Dolmuşlar} \sum_{i=0}^{\#Dolmuşlar} I_2(T_g, Dolmuş_i) \quad (7)$$

$$P_1 = C_1 - V_1 \quad (8)$$

Yeni bir geçerli tespit (T_g) için P_1 , eğitim veri setindeki binek araçlara olan ortalama l_2 (9) uzaklığı ile eğitim veri setindeki dolmuşlara olan ortalama l_2 uzaklık farkıdır. Sözü edilen l_2 uzaklığı 7 Hu momentine [15] göre aşağıdaki gibi hesaplanır:

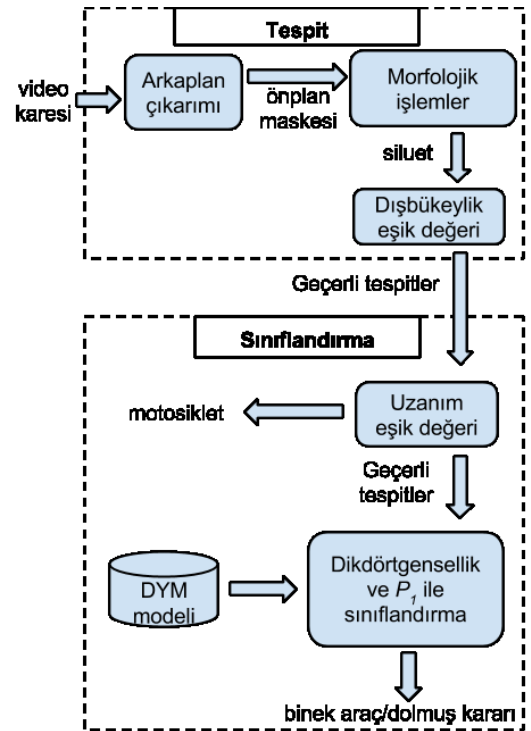
$$I_2(A, B) = \sum_{i=1..7} |m_i^A - m_i^B| \quad (9)$$

$$m_i^A = işaret(h_i^A) \cdot \log h_i^A \quad (10)$$

$$m_i^B = işaret(h_i^B) \cdot \log h_i^B \quad (11)$$

Burada h_i^A ve h_i^B A ve B şekillerinin Hu momentleridir [16].

Eğer bir tespit, motosiklet olarak sınıflandırılmadıysa, ya binek araç ya dolmuş olarak sınıflandırılır. Dikdörtgensellik ve P_1 özneliklerini kullanarak binek araç ve dolmuş sınıfları arasında doğrusal bir karar sınırı çizilebilir. Bu sınır eğitim kümesindeki örnekler ve DYM (Destek Yöney Makinesi) kullanarak belirlenmiştir.

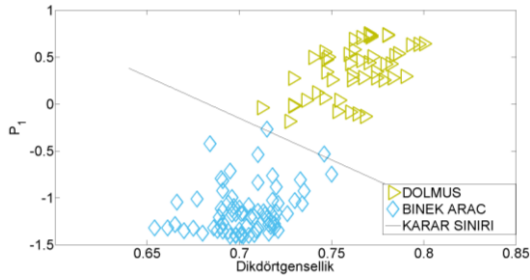


Şekil 4. Sistemin akış şeması. Tek bir video karesi kullanan yöntemde tek bir silüet, önerilen yöntemde ise ortalama silüet kullanılmıştır.

IV. DENEYLER

Canon 600D SLR kamera ve bir aynalı aparat (www.gopano.com) kullanarak oluşturduğumuz tümyönlü kamera ile 49 motosiklet, 124 binek araç ve 104 dolmuştan oluşan toplamda 277 taşıt videolu veri seti elde ettik. Veri setini eğitim ve test olmak üzere ikiye ayırdık. Eğitim seti, tüm veri setinin %60'ını kapsamaktadır.

Tüm deneylerde $\rho=0.75$ sabit tutulmuştur. Uzanım eşik değeri ise, motosiklet ve diğer araçları ayırdığı için, eğitim setindeki motosikletlerin maksimum uzanım değeri olarak belirlenmiştir. Dikdörtgensellik ve P_1 öznelikleri ile eğitilen binek araç/dolmuş sınıflandırıcısının bir örneği Şekil 5'te verilmiştir.



Şekil 5: Ortalama silüet ile eğitilen DYM modeli.

Ortalama silüet metodu, eğitim setinden çıkan ortalama silüetler ile eğitilmiştir. Tek silüet metodunu tek bir video karesinden elde edilen silüetlerle eğitmek, bu kareler bozulmuş silüetler içerdiği için, adil olmayacağından bu yöntem için taşıtların sınırları el ile etiketlenmiştir.

Örneklerin rastgele dağıldığına emin olmak için işlemler veri setini eğitim ve test olmak üzere farklı şekilde bölerek üç kere tekrarlanmıştır. Karşılaştırılan iki yöntem için üç deneyin ortalaması Tablo 1'de doğru sınıflandırılan taşıt yüzdesi, Tablo 2 ve 3'te ise her bir yöntem için hata matrisi olarak verilmiştir. Ortalama silüet yöntemi daha iyi bir performans sunmaktadır. Tablolarda belirtilen yanlış eksiler, dışbükeylik eşik değeri ile elenen geçerli sayılmayan fakat gerçek silüetlerdir. Şekil 6, ortalama silüet yöntemi ile doğru, fakat tek silüet yöntemi ile yanlış sınıflandırılan bir örneği göstermektedir.

Tablo 1. Her bir sınıf için doğru sınıflandırılan taşıt yüzdesi

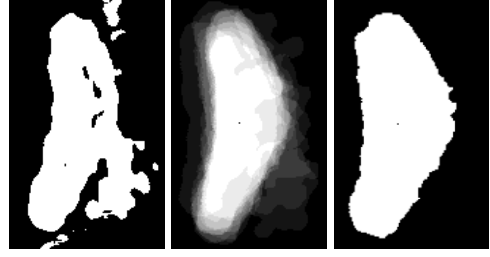
	Motosiklet	Binek araç	Dolmuş	Toplam
Ortalama silüet	95%	98%	83%	92%
Tek silüet	80%	78%	81%	79%

Tablo 2. Tek silüet yöntemi için hata matrisi

Gerçek sınıf		Motosiklet	Binek araç	Dolmuş
Tespit	Motosiklet	16	3	4
	Binek araç	0	39	1
	Dolmuş	1	7	34
	Yanlış-eksi	3	1	3

Tablo 3. Ortalama silüet kullanan önerilen yöntemin hata matrisi

Gerçek sınıf		Motosiklet	Binek araç	Dolmuş
Tespit	Motosiklet	19	0	0
	Binek araç	0	49	1
	Dolmuş	1	1	35
	Yanlış-eksi	0	0	6



Şekil 6: Binek araç örneği. Solda yanlış sınıflandırılan tek silüet. Ortada ortalama silüet. Sağda eşik değeriyle aydınlatılmış ortalama silüet.

V. VARGILAR

Hareketli nesneye ait silüetleri kullanan şekil tabanlı bir taşıt sınıflandırma yöntemi öne sürdük. Yöntemimizi hem tek video karesinden çıkan silüete, hem de çok sayıda karenin zamansal ortalaması olan silüete uyguladık. Ortalama silüet ile yapılan sınıflandırmanın tek silüet ile yapılandan daha başarılı olduğunu deneylerde gördük. Ayrıca, metodumuz benzerlerinin aksine kamera-nesne arası uzaklığa bağlı bir öznelik kullanmadığından taşınabilir görüntü alma platformları için uygundur.

KAYNAKÇA

- [1] Dalal, N. and Triggs, B., "Histograms of Oriented Gradients for Human Detection", CVPR, 2005.
- [2] Cinaroglu, I. and Bastanlar, Y., "A direct approach for object detection with catadioptric omnidirectional cameras", Signal, Image and Video Processing, DOI: 10.1007/s11760-015-0768-2, 2015.
- [3] Karaimer, H.C. and Bastanlar, Y., "Car detection with omnidirectional cameras using haar-like features and cascaded boosting", Signal Proces. Comm. Applications Conf. (SIU), 2014.
- [4] Dupuis, Y., Savatier, X., Ertaud, J., Vasseur, P., "A direct approach for face detection on omnidirectional images", ROSE, 2011.
- [5] Morris, B. and Trivedi, M., "Improved vehicle classification in long traffic video by cooperating tracker and classifier modules", AVSS, 2006.
- [6] Gupte, S., Masoud, O., Martin, R., and Papanikolopoulos, N., "Detection and classification of vehicles", Intelligent Transportation Systems, 37-47, 2002.
- [7] Kumar, P., Ranganath, S., Weimin, H., and Sengupta, K., "Framework for real-time behavior interpretation from traffic video", Intelligent Transportation Systems, 43-53, 2005.
- [8] Buch, N., Orwell, J., and Velastin, S., "Detection and classification of vehicles for urban traffic scenes", Visual Information Engineering, 182-187, 2008.
- [9] Mithun, N., Rashid, N., and Rahman, S., "Detection and classification of vehicles from video using multiple time-spatial images", Intelligent Transportation Systems, 1215-1225, 2012.
- [10] Khoshabeh, R., Gandhi, T., and Trivedi, M., "Multicamera based traffic flow characterization and classification", ITSC, 2007.
- [11] Gandhi, T. and Trivedi, M., "Video based surround vehicle detection, classification and logging from moving platforms: Issues and approaches", 1067-1071, 2007.
- [12] Sobral, A. and Vacavant, A., "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos", Computer Vision and Image Understanding, 2014.
- [13] Yao, J. and Odobez, J., "Multi-layer background subtraction based on color and texture", CVPR, 1-8, 2007.
- [14] Yang, M., Kpalma, K., Ronsin, J., "A survey of shape feature extraction techniques", Pattern recognition, 43-90, 2008.
- [15] Hu, M.-K. "Visual pattern recognition by moment invariants.", Information Theory, IRE Transactions, 179-187, 1962.
- [16] Bradski, G. and Kaehler, A. "Learning OpenCV: Computer Vision with the OpenCV Library", O'Reilly Media, 2008.