Kafa Duruşu Kestirimlerinden Bakış Yönünün Türetilmesi
Derivation of Gaze Direction from Head Pose Estimates

Zeynep Yücel [⋆], *Albert Ali Salah*[†], *Çetin Meriçli*[*‡], *Tekin Meriçli*[‡]

[⋆] IRS Laboratuvarı, ATR International, Kyoto
[†] ISLA Laboratuvarı, Amsterdam Üniversitesi, Amsterdam
[*] Coral Araştırma Grubu, Carnegie Mellon Üniversitesi, Pittsburgh
[‡] AI Laboratuvarı, Boğaziçi Üniversitesi, İstanbul
zeynep@atr.jp, a.a.salah@uva.nl, cetin@cmu.edu, tekin.mericli@boun.edu.tr

Özetçe

Bakış yönü bilgisi özellikle insan-robot ve insan-bilgisayar etkileşimi uygulamaları için çok önemli bir bilgidir. Uygulamanın niteliğine bağlı olarak bu bilgiyi düşük çözünürlükü görsel girdiden gerçek zamanlı ve yüksek kesinlikle elde etmek gerekebilir. Bu çalışmada bu gibi durumlarda kullanılabilecek bir bakış yönü çözümleme yöntemi önerilmektedir. Bu çalışmanın esas katkısı kafa duruşu ve bakış yönü arasında kesin bir ayrım yapmasında yatmaktadır. Bu alanda daha önce yapılmış bazı çalışmalarda olduğu gibi kafa duruşu üzerinde deney düzenlemesine göre uygun bir şekilde düzeltme yapmak yerine çevreden bağımsız bir dönüşüm uygulanması önerilmektedir. Tanımlanan yöntemle eliptik silindir tabanlı kafa duruşu çözümesiyle elde edilmiş kestirimler bakış yönüne dönüştürülmektedir. Bu dönüşüm için bir Gauss süreci bağlanımı kullanılması önerilmekte ve bu önermenin geçerliliği ayrıntılı bir şekilde tartışılmaktadır.

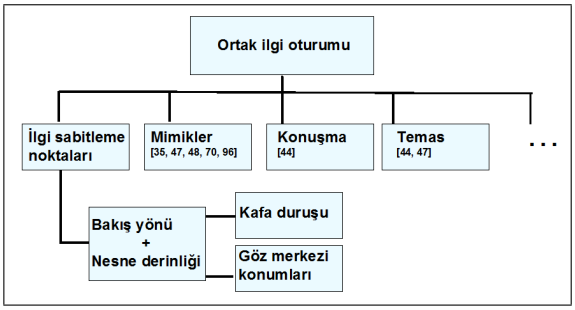
Abstract

Gaze direction information is very important in particular for certain applications of human-robot and human-computer interaction. Depending the properties of the specific application, it may be required to derive this information in real time from low resolution visual inputs with high precision. In this paper an algorithm suitable for such cases is described to transform the head pose estimates to gaze direction. The main contribution of this study lies in the fact that it makes a clear distinction between head pose estimates and gaze direction. Unlike some of the previous works in this field, we do not correct the head pose to correspond to a possible attention fixation point in accordance with the experiment scenario. Instead we propose using a concrete and environment-independent method for this purpose. To transform the head pose estimates into gaze direction, a Gaussian process regression model is proposed to be employed and the reasons validating this choice are discussed in detail.

1. Giriş ve İlgili Çalışmalar

[1, 2, 3] Bu çalışmada ortak ilgi oturumunun kurulması problemine eğilinerek bu amaç için kullanılabilecek bir bakış yönü bulma yöntemi önermekteyiz. Kullanılan video dizilerinin çözünürlüğünden dolayı ilgi sabitleme noktalarının iki önemli belirleyicisi olan göz merkezi konumları ve kafa duruşundan sadece kafa duruşunu kullanmayı seçmiş bulunmaktayız. Ancak kafa duruşunun ilgi sabitleme noktalarını doğrudan ve tek başına tanımlamaya yetmeyeceğini de önemle vurgulamaktayız. Bu sebepele Gauss süreci bağdaşımı ile bir dönüşüm yapmayı önermekteyiz.

Bu çalışmanın akışı şu şekildedir: Bölüm 2'de veri tabanı ve deney düzenlenişinin ayrıntıları sunulmaktadır. Bölüm 3 önerilen yöntemin detaylarını açıklamaktadır. Gauss süreci bağlanımı Bölüm 3.1'de ele alınırken, model seçimi Bölüm 3.2'de açıklanmıştır. Son olarak Bölüm 4'te önerilen yöntemin katkıları ve çalışmanın sonuçları özetlenmektedir.



Şekil 1: Ortak ilgi oturumunun temel bileşenleri.

2. Veri Tabanlı ve Deneyler

Bakış yönünü belirlemek için Boğaziçi Üniversitesi Yapay Zeka Laboratuvarında oluşturulmuş bir deney kümesinden faydalanılmıştır. Söz konusu deney kümesi hareketli bir robot üzerine yerleştirilmiş bir kamera tarafından 10 fps çerçeve hızı (frame rate) ile 4 farklı kişiden kaydedilmiş toplam 8 video dizisinden oluşmaktadır.

Deneylerin düzenlenişi şu şekildedir. Deneyi yapan kişi öncelikle robotun üzerindeki kameraya bakarak göz teması sağlamakta ve ortak ilgi oturumunun başladığı işaretini vermektedir. Ardından önündeki masa üzerinde bulunan 6 değişik objenin her birine en az bir kez belirli bir süre boyunca rastlantısal

bir sıra ile bakmaktadır. Bu sırada robotun üzerinde bulunan kamera bu sahneyi kaydetmektedir. Kaydedilen görüntüye eliptik silindir tabanlı kafa duruşu çözümleme algoritması uygulanmakta ve her video çerçevesine tekabül eden kafa duruşu hesaplanmaktadır. Eliptik silindir tabanalı kafa duruşu çözümleme algoritmasının ayrıntıları ve bazı uygulamaları önceki yayınlarımızda bildirilmiştir. Her çerçeve için kafa duruşu çözümlemesi yapıldıktan sonra Bölüm 3'de ayrıntıları verilen algoritma izlenerek çözümlenen kafa duruşlarından bakış yönü türetilmektedir.

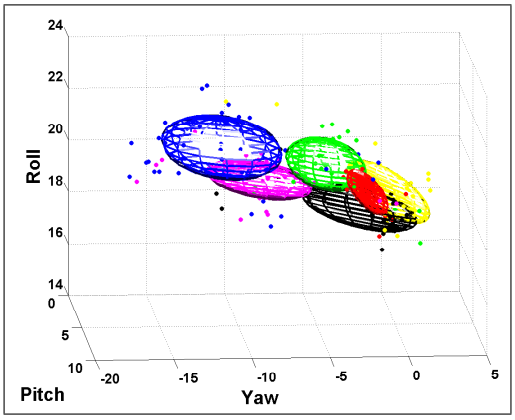
3. Yöntem Bilgisi

Bakış yönünün belirlenmesinde kafa duruşu ve gözlerin konumu iki önemli bir etmendir. Bakış yönünün tesbitinde en sa bu iki etmenin bileşeninin ele alınması gerekmektedir. Ne var ki ele aldığımız veri tabanı düşük çözünürlüklü video dizilerinden oluşmaktadır. Bu sebeple kafa duruşunu belirlemek gözlerin konumunu belilemeye kıyasla daha kolaydır. Hatta Şekil 2'de görüldüğü gibi göz bölgeleri oldukça az sayıda pikselden oluştuğudan göz merkezlerinin konumu sağlıklı bir şekilde hesaplamak oldukça zordur.

Şekil 2: Yaklaşık 15×25 pikselden oluşan göz bölgeleri.

Bu sebeple bakış yönünün bulunmasında sadece kafa duruşunu kullanmayı öneriyoruz. Ancak kafa duruşu ve bakış yönünün yakından ilişkili olmalarına rağmen aynı olmadıklarını da bir kez daha vurguluyoruz. Bu sebeple kafa duruşunu bakış yönüne eşit kabul etmek yerine bu ikisi arasında bir dönüşüm tanımlamayı öneriyoruz.

Tanımlanacak dönüşümün modelini belirlemek içinse kafa duruşu verilerine daha yakından bakmak ve elimizdeki değişkenlerin özelliklerini daha iyi anlamak gerekmektedir. Öncelikle kafa duruşu verilerine bakalım. Şekil 3 örnek video dizisi için kafa duruşu değerlerinin dağılımını vermektedir. Burada duruş açılarının demetler (topaklar, cluster) oluşturduğu açıkça görülmektedir. Ayrıca [3]'deki deney düzeneği incelendiğinde duruş açıları ile odaklanılan nesne yerleşiminin doğrusal olmayan bir ilişki içinde olduğu da belirlenmektedir. Bu gözlem kafa duruşunun bakış yönünün önemli bir gösterdegesi olduğu iddiasını desteklemektedir. Bugüne kadar bazı çalışmalar bir takım ek varsayımlarda bulunarak bu problemin üstesinden gelmeye çalışmışlardır. Örneğin [4]'de yazarlar ilgi odağının bir insan üzerinde oldugunu varsaymış ve çözümlenen kafa duruşunu bu varsayıma göre düzeltmişlerdir. Ancak deney düzenlenişindeki ve amaçlardaki farklardan dolayı böylesi bir düzeltme yöntemi bizim problemimize uygulanamaz. Bu sebeple bu sorunu doğrusal olmayan bir ifade benimseyerek aşmayı öneriyoruz. Bu amaçla da bir Gauss süreci bağlanımı kullanmayı teklif ediyoruz. Bu yolla



Şekil 3: Örnek bir video dizisi için kafa duruşunun dağılımı.

kestirilmiş kafa duruşu değerlerinden bakış yönünü çevreye göre değişen bir yöntem yerine kurallı bir çözümleme ile türetilebileceğini öneriyoruz.

3.1. Gauss süreci bağlanımı

Varsayalım ki x bir duruş yöneyi, y ise x'e karşılık gelen sayıl (scalar) hedef değer olsun. Bu problemin ele alınışına göre bu hedef değer bakış yönü olacaktır. Amacımız x ve y arasındaki ilişkinin özelliklerini incelemektir.

Varsayalım ki y hedef değerleri x kafa duruşu kestirimlerinden $f(.)$ gibi bir dönüşüm kullanılarak elde ediliyor olsun.

$$y = f(x). \tag{1}$$

Dönüşüm fonksiyonu f Gauss süreci olduğu bilinen bir dağılımdan gelsin. Bu durumda f, ortalama değer fonksiyonu $m(x)$ ve kovaryans fonsiyonu $k(x, x')$ ile tamamen tanımlanabilmektedir.

$$f(x) \sim \mathcal{G}P(m(x), k(x, x')). \tag{2}$$

Beklenen değer $E[.]$ ile gösterildiğinde [5], ortalama değer fonksiyonu ve kovaryans fonksiyonu aşağıdaki şekilde tanımlanır:

$$\begin{aligned} m(x) &= E[f(x)], \\ k(x, x') &= E[(f(x) - m(x))(f(x') - m(x'))]. \end{aligned} \tag{3}$$

Bu aşamadan sonra ifadelerde sadelik sağlamak amacıyla $m(x) = 0$ olduğunu varsayıyoruz. Bu varsayım bir göreli kaydırma (offset) uygulayarak kolayca telafi (compensate) edilebilir.

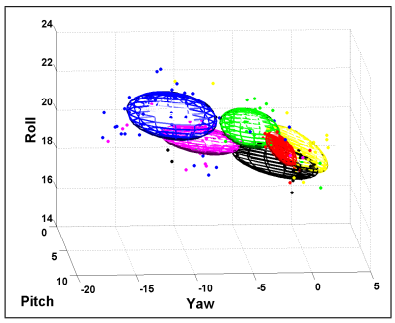
Gauss süreci bağdaşımı varsayımına ek olarak değişkenlerin gözlemlenen değerlerinin de gürültülü olduğu varsayılmaktadır. Bu ele alınan problemin formülizasyonunu gerçek hayat uygulamalarına daha yakın hale getirmektedir. Aynı zamanda bu varsayım göz hareketlerinin bakış yönü üzerindeki etkisini hesaba katmak olarak da görülebilir. Gözlemlenen değerlerin gürültülü olduğu ve Gauss süreci bağdaşımından geldiği varsayımlarını aşağıdaki savunma (argument) yoluyla doğrulayacağız. (validate)

Bakış doğrutusunu $\vec{e}_d$ ile, göz merkezini $e_0 = [x_0, y_0, z_0]$ ile, iris merkezini de $i_0 = [x_i, y_i, z_i]$ ile gösterelim. Göz yuvarlağı bir küre olarak modellendiğinde, bakış doğrultusu $\vec{e}_d$, e_0 noktasını i_0 noktasına bağlayan yöney olarak tanımlanır. Bu durumda, O başlangıç noktasını (origin) gösteriyorken,

$$\vec{e}_d = \overline{Oi_0} - \overline{Oe_0}, \tag{4}$$

olur. Kişinin belirli bir anda odaklandığı nokta ise gözün bakış doğrultusu $\vec{e}_d$ yöneyi ile hedef düzlemin kesişimi hesaplanarak bulunur.

Varsayalım ki deneyi yapan kişi Şekil 4'de gösterilen hedef bölgeye bakıyor olsun. Hedef bölge üzerinde kişinin odaklandığı noktanın konaçlarının (coordinate) da $a_0 = (x_{a_0}, y_{a_0})$ şeklinde olduğunu düşünelim. Bu şekilde x ekseni etrafındaki



Şekil 4: Örnek bir video dizisi için kafa duruşunun dağılımı.

dönüşü θ, y ekseni etrafındaki dönüşü ϕ, z ekseni etrafındaki dönüşü ise ψ ile gösterelim. Bu açıları dağılımlarının *bağımsız ve özdeş dağılmış* (independent and identically distributed) olduğunu göstermek istiyoruz.

Kişinin baktığı nokta a_0 ile ilgili olarak sadece x_{a_0} konacının bilindiğini düşünelim. Bu bilgi bize y_{a_0} konacıyla ilgili bir bilgi sunmayacaktır.

$$p(x_{a_0}|y_{a_0}) = p(x_{a_0}). \tag{5}$$

Aynı şekilde y_{a_0} konacının bilinmesi de x_{a_0} konaçının çözümlenmesinde fayda sağlamayacaktır. Bu sebeple x_{a_0} ve y_{a_0} konaçlarının bağımsız olduğunu öne sürebiliriz. Böylece

$$\mathrm{cov}(x_{a_0}, y_{a_0}) = 0 \tag{6}$$

olur. Öte yandan, kişinin hedef tahtası üzerindeki bütün noktalara bakma olasılığı da eşit kabul edilmektedir:

$$\begin{aligned} p(x_{a_0}) &= \frac{1}{(x_f - x_i)},\ x_{a_0} \in [x_i, x_f], \\ p(y_{a_0}) &= \frac{1}{(y_f - y_i)},\ y_{a_0} \in [y_i, y_f]. \end{aligned} \tag{7}$$

Burada Hoffman vd.'nin [6]'de yaptığı gibi olasılıksal bir yaklaşım benimsemeyi uygun bulmuyoruz. Bu varsayım özellikle robot öğrenmesi gibi uygulamalarda yanılmalara yol açabilir. Bu gibi durumlarda kişinin bakış tercihlerini şartlandırmamak daha uygundur. Bu sebeple özdeş dağılım özelliğini tercih ediyoruz.

Hedef tahtası üzerindeki noktaların bakış tercihine göre özelliklerini bu şekilde belirledikten sonra bunu bakış yöüne doğru geriye götüreceğiz. Burada 3B (3 boyutlu) göz yuvarlağı uzayından 2B hedef tahtası uzayına doğrusal bir haritalama olduğunu da öne sürüyoruz. !!

Bu ifade aşağıda görülen Denklem 8 vasıtasıyla formüle edilebilir:

$$y = f(x) + \varepsilon_0, \tag{8}$$

Burada ε_0 bağımsız ve σ_0^2 varyansı ile identically dağılmış beyaz gürültüyü göstermektedir. Varsayalım ki n gözlem ikilisi içeren $\mathcal{D}$ gibi bir eğitim kümemiz olsun,

$$\mathcal{D} = \{(x_i, y_i) | 1 \leq i \leq n\}. \tag{9}$$

Girdiler X gibi bir matriste ve hedef değerler de $\mathbf{y}$ gibi bir yöneyde toplandığında, $\mathcal{D}$ eğitim kümesinin içerdiği toplam bilgi şu şekilde ifade edilebilir:

$$\mathcal{D} = (X, \mathbf{y}). \tag{10}$$

Gözlem kümesi, $\mathcal{D}$ Gauss süreci bağlanımının *eğitiminde* kullanılmaktadır. Bu model seçimi ve parametre optimizasyonunun $\mathcal{D}$ tarafından sağlanan bilgi kullanılarak yapıldığı anlamına gelmektedir. Bunun icin de hedef değerlerin şartlandırılmış (conditional) dağılımının bulunmasında bir Bayesian yöntem uyguluyoruz.

Varsayalım ki X girdi matrisi n tane eğitim örneğinden oluşsun ve $X^\star$ test matrisi de da $n^\star$ tane test noktasından oluşsun. Buna göre $\mathbf{f}^\star$ dönüşümü aşağıdaki şekildir

$$\mathbf{f}^\star \sim \mathcal{N}(0, K(X^\star, X^\star)). \tag{11}$$

Burada K covariance matrisini belirtmektedir. Eğitim çıktıları olan $\mathbf{y}$'lerin ve test çıktıları olan $\mathbf{f}^\star$'lerin öncül (prior) bileşik dağılımı şu şekilde yazılabilir:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}^\star \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma_0^2 I & K(X, X^\star) \\ K(X^\star, X) & K(X^\star, X^\star) \end{bmatrix}\right). \tag{12}$$

Artçıl (posterior) dağılım da öncül dağılım sadece gözlemlerle tutarlı olan fonksiyonları içercek şekilde sınırlandırılarak elde edilir. Bu sebeple öncül dağılım gözlemlerle şartlandırıldığında, aşağıdaki ifadeyi elde ederiz:

$$\mathbf{f}^\star | X, \mathbf{y}, X^\star \sim \mathcal{N}(\bar{\mathbf{f}}^\star, \mathrm{cov}(\mathbf{f}^\star)), \tag{13}$$

Burada

$$\begin{aligned} \bar{\mathbf{f}}^\star &= K(X^\star, X)[K(X,X) + \sigma_0^2 I]^{-1}\mathbf{y}, \\ \mathrm{cov}(\mathbf{f}^\star) &= K(X^\star, X^\star) - \\ &\quad K(X^\star, X)[K(X,X) + \sigma_0^2 I]^{-1} K(X, X^\star). \end{aligned} \tag{14}$$

Ortalama kestirimler gözlemlerin doğrusal birleşimi (linear combination) olduğu için, Denklem 14 zaman zaman *doğrusal kestirici* olarak da isimlendirilir [5].

3.2. Model seçimi

Bölüm 3.1'de de bahsedildiği gibi, amacımız x ve y arasındaki illişkinin ayrıntılarını ortaya çıkarmaktır. Bu da Gauss süreci f'nin yapısını anlamayı gerektirir. Bu bakımdan hiyerarşik bir yaklaşım uygulamak sıkça benimsenen bir tutumdur. Modelin en üset seviyesinde f'nin ait olduğu modelin yapısı $\mathcal{H}$'nın olduğunu varsayalım. Bir seviye aşağıda ise model

parametrelerinin daŭılımını gösteren θ hiperparametreleri olsun. Model parametreleri ise w ile gösterilmektedir ve en alt seviyede yer almaktadırlar.

Model seçimi terimi model yapısı $\mathcal{H}$'nin seçimi gibi ayrık seçimleri ifade edebileceği gibi θ hiperparametrelerinin optimizasyonunu da kapsar. Bunun sebebi bu problemlerin aynı biçimde, yani Bayesian bir yaklaşım kullanarak çözülmesidir.

Bayes kuralına göre en al seviyede bulunan artçıl dağılımı olabilirlik (likelihood), marjinal olabilirlik (marginal likelihood) ve öncül dağılım cinsinden şu şekilde buluruz:

$$p(\mathbf{w}|\mathbf{y}, X, \theta, \mathcal{H}_i) = \frac{p(\mathbf{y}|X, \mathbf{w}, \mathcal{H}_i)p(\mathbf{w}|\theta, \mathcal{H}_i)}{p(\mathbf{y}|X, \theta, \mathcal{H}_i)}. \quad (15)$$

Burada $p(\mathbf{y}|X, \mathbf{w}, \mathcal{H}_i)$ olabilirlik terimidir. öncül dağılım $p(\mathbf{w}|\theta, \mathcal{H}_i)$ ile gösterilir. Kanıt terimi de denen marjinal olabilirlik ise

$$p(\mathbf{y}|X, \theta, \mathcal{H}_i) = \int p(\mathbf{y}|X, \mathbf{w}, \mathcal{H}_i)p(\mathbf{w}|\theta, \mathcal{H}_i)d\mathbf{w}. \quad (16)$$

Hiyerarşik formulasyonun bir sonraki aşaması ise şu şekildedir:

$$p(\theta|\mathbf{y}, X, \mathcal{H}_i) = \frac{p(\mathbf{y}|X, \theta, \mathcal{H}_i)p(\theta|\mathcal{H}_i)}{p(\mathbf{y}|X, \mathcal{H}_i)}. \quad (17)$$

Burada

$$p(\mathbf{y}|X, \mathcal{H}_i) = \int p(\mathbf{y}|X, \theta, \mathcal{H}_i)p(\theta|\mathcal{H}_i)d\theta. \quad (18)$$

En üst seviyede ise aşağıdaki ifade yer alır.

$$p(\mathcal{H}_i|\mathbf{y}, X) = \frac{p(\mathbf{y}|X, \mathcal{H}_i)p(\mathcal{H}_i)}{p(\mathbf{y}|X)}, \quad (19)$$

Burada marjinal olabilirlik aşugıda belirildiği gibidir.

$$p(\mathbf{y}|X) = \sum_i p(\mathbf{y}|X, \mathcal{H}_i)p(\mathcal{H}_i). \quad (20)$$

4. Sonuçlar

Bu çalışmada kafa duruşu kestirimlerinden ilgi sabitleme noktalarının türetilmesini sağlayacak bir yöntem tanımlanmıştır. Özellikle göz merkezi konumlarını bulmanın zor olduğu düşük çözünürlüklü video dizilerinde fayda sağlayacak bu yöntem bakış yönünün Gauss süreci olduğu bilinen bir dönüşüm ile kafa duruşu kestirimlerine çevrilebileceği gerçeğinden faydalanır. Bu varsayımın geçerliliği ayrıntılı bir şekilde tartışılmıştır. Bu çalışmanın en önemli katkısı

the restrictions introduced by the database of interest, the information obtained by an elliptic cylindrical head model based pose estimator is transformed into gaze direction. It is proposed to use a Gaussian process regression model for this transformation and the reasons validating this choice are discussed in detail. The details of Gaussian process regression and model selection are explained in particular. The main contribution of this study lies in the fact that it makes a clear distinction between head pose estimates and gaze direction. Unlike some of the previous works in this field, we do not correct the head pose to coincide with a possible attention fixation point with a reasonable scheme according to the experiment scenario. Instead we define a concrete and environment-independent method for this purpose.

Teşekkür: Bu çalışma TÜBİTAK 107A011 nolu proje desteği ve Devlet Planlama Teşkilatı proje desteği ile gerçekleştirilmiştir.

5. Kaynakça

[1] Z. Yücel and A. A. Salah. Head pose and neural network based gaze direction estimation for joint attention modeling in embodied agents. In *Proceedings of the Annual Meeting of Cognitive Science Society*, pages 3139–3144, 2009.

[2] Z. Yücel and A.A. Salah. Resolution of focus of attention using gaze direction estimation and saliency computation. In *Proceedings of the International Conference on Affective Computing and Intelligent Interfaces*, 2009.

[3] Z. Yücel, A.A. Salah, C. Meriçli, and T. Meriçli. Joint Visual Attention Modeling for Naturally Interacting Robotic Agents. In *Proceedings of the* 24^{th} *International Symposium on Computer and Information Sciences*, pages 242–247, 2009.

[4] R. Stiefelhagen, J. Yang, and A. Waibel. Modeling focus of attention for meeting indexing. In *Proceedings of the* 7^{th} *ACM international conference on Multimedia (Part 1)*, pages 3–10, 1999.

[5] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[6] M.W. Hoffman, D.B. Grimes, A.P. Shon, and R.P.N. Rao. A probabilistic model of gaze imitation and shared attention. *Neural Networks*, 19(3):299–310, 2006.