Semi-supervised Co-update of Multiple Matchers

Luca Didaci[1], Gian Luca Marcialis[2], and Fabio Roli[2]

University of Cagliari
[1] Department of Pedagogical and Philosophical Sciences
[2] Department of Electrical and Electronic Engineering
Piazza d'Armi - 09213 Cagliari, Italy
{luca.didaci,marcialis,roli}@diee.unica.it

Abstract. Classification algorithms based on template matching are used in many applications (e.g., face recognition). Performances of template matching classifiers are obviously affected by the representativeness of available templates. In many real applications, such representativeness can substantially decrease over the time (e.g., due to "aging" effects in biometric applications). Among algorithms which have been recently proposed to deal with such issue, the template co-update algorithm uses the mutual help of two complementary template matchers to update the templates over the time in a semi-supervised way [8]. However, it can be shown that the template co-update algorithm derives from a more general framework which supports the use of more than two template matching classifiers. The aim of this paper is to point out this fact and propose the co-update of multiple matchers. Preliminary experimental results are shown to validate the proposed model.

1 Introduction

Template matching is a widely used classification approach in many applications [1-3]. For example, in personal verification systems based on multi-modal biometrics [2-4]. However, the effectiveness of these approaches is strongly dependent on the "representativeness" of templates. For example, in biometric applications, intra-class variations and aging (scratch on the finger skin or the face, unknown expressions, etc.) may affect such representativeness [2-3]. In order to deal with these problems, several "template update" algorithms have been proposed. Templates are usually captured under the supervision of a human expert (e.g. during the enrolment phase for biometric verification systems), but capturing all possible intra-class variations is impossible.

With regard to this issue, "semi-supervised" template update has been proposed [5-8]. In this approach, unlabelled samples (that is, samples collected during the system's operation) showing very strong similarity to existing templates are added into the set of samples of the related class, thus increasing the representativeness of related set of templates. Recently, the so-called "co-update" algorithm, derived from the co-training method [9], showed that the mutual "help" of two weakly correlated matchers can increase the template representativeness more quickly than systems adopting only one matcher, and improve the overall performance. This has been verified, in particular, for biometric applications [7-8].

J.A. Benediktsson, J. Kittler, and F. Roli (Eds.): MCS 2009, LNCS 5519, pp. 152–160, 2009.

In [7], a statistical framework modelling the co-update behaviour has been proposed in the biometric scenario. This model lies on hypotheses which fit the practical scenarios involving fingerprints and faces, providing a first analytical framework able to explain the relationship between the amount of unlabelled data collected and the template set size for a certain application. However, it can be shown that this model derives from a general framework which involves more than two matchers, independently on the peculiar application. The aim of the present work is to point out this fact by proposing the "co-update of multiple matchers". With this term, we mean a generic classification system based on template-matching [1]. Preliminary experiments are done on a case-study of a bi-modal biometric verification system. Results show the reliability of the framework by comparing the template set sizes and errors predicted by the model with the template set sizes and errors obtained on test data.

The paper is organized as follows. Section 2 describes the current version of the model. Section 3 reports the experimental results which validate the model, and also shows the difference with the previous one. Section 4 concludes the paper.

2 Co-updating of Multiple Matchers

Let us consider a pattern recognition problem characterized by c sets of features (here called "views"), and each view is conditionally independent of each other. The system is made up of c matchers, one matcher for each view. Thanks to the independence assumption, each recognizer is expected to assign correct "labels" to certain input data which are difficult for the other. In the case of template co-update, two matchers are used (bi-modal system) [7, 8]. In this Section, we show that the model proposed in [7] derives from a general framework involving more than two matchers.

The description of the generalized co-update algorithm is given in Figure 1. Let D_l be a set of labelled data. These data are the templates captured by human supervision, thus they are labelled with classes identifiers (e.g. users names in the case of personal verification). It is supposed that a batch D_u (usually, much larger than D_l) of data is acquired during system operation. Each element of D_u is a c-uple of samples, one sample for each view. Matchers can be combined in different kinds of ensembles (abstract-level, score-level [10]). Without loss of generality, we can neglect the particular kind of adopted fusion rule. By setting the size of the ensemble Dim_ens, $N\_ens = \binom{c}{Dim\_ens}$, different ensembles are possible. The main idea of co-updating of multiple matchers is that each ensemble can update the template set of remaining c-Dim_ens matchers. In this paper, we adopted the same terminology used in [7] for sake of uniformity with the state-of-the-art. Let us call "master" the ensemble that assume the supervisor's role, and "slave" the remaining matchers, whose template sets are augmented thanks to the master ensemble.

During the off-line co-update phase, each ensemble is applied to the batch D_u. The results of matching is a value which can be interpreted as a similarity value between the input sample and related template set. If this value exceeds a given "updating threshold", the match is verified, and 'slave' samples in the c-uple are added to the set

D_l. All matchers assume, alternatively, master and slave roles. The process is repeated a specified number of times.

It is worth highlight two particular cases. (1) *Dim_ens*=1: the ensemble is made up of only one matcher at time, the other ones are 'slaves'. (2) *Dim_ens*=*c*-1, the ensemble is made up of c-1 matchers at time, and the remaining one is 'slave'.

```
Define $c$ matchers, a matcher for each feature set x1, …, xc
Set the size of the ensemble $Dim\_ens$
Set the number of possible ensembles $N\_ens = \binom{c}{Dim\_ens}$

In the enrolment (supervised) session, collect a set $D_l$ of samples (in terms of c feature sets). A c-uple of samples $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_C\}$ – one sample for each matcher- is acquired.
Create templates using the set $D_l$
Loop for $H$ iterations:
    Collect a set $D_u$ without supervision. Each element of $D_u$ is a c-uple of samples $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_C\}$ from the same class
    For each ensemble $E_i$, $i \in \{1, 2, ..., N\_ens\}$
        Let the matchers used by $E_i$ be the 'master', while the others are the 'slaves'
        For each c-uple of elements in $D_u$
            If the ensemble $E_i$ (master) verifies the matching (i.e. match exceed a pre-defined updating threshold), the samples in the c-uple related to 'slave' matchers are added to the set $D_l$
Update templates using the augmented labelled set $D_l$
```

Fig. 1. Algorithm for Co-updating of multiple matchers

As for the majority of semi-supervised template update algorithms, the core of co-update is the insertion of new samples without external supervision. This might introduce misclassified samples in template set. For this reason, only "confident examples" are added by a very high co-update threshold value at which no false matches occur, namely, FAR = 0%. The role, master or slave, of the matchers is highlighted by superscripts M and S. For sake of clarity, we omit in the following terminology the explicit reference of each slave matcher.

Some assumptions are made in order to propose a mathematically tractable model. The first assumption regards the finite discrete space used to represent different possible values for each feature set (for example, a certain class can be represented by a finite number of possible directed acyclic graphs). Each class can produce a maximum of $N_{TOT}^{(i)}$ different observations for each of the *c* representations at hand. This assumption can be motivated easily noting that only a finite number of samples are collectable in a physical system. The main consequence of this assumption is that a new example, very similar to an existing one, can be considered already present in the template set.

The second assumption states that all possible observations are equally probable (e.g. this is usually stated in biometric applications), and the c representations are conditionally independent given the class [8,9]. In the co-update stage, a set D_u of $|D_u|=k$ c-uples of samples are presented to the system. We indicate with k' the number of samples of the master ensemble whose class has been verified (i.e. match exceeds the given updating threshold). The value k' can be easily computed by considering that each of k samples of the master ensemble can be drawn with replacement from a homogeneous population of $N_{TOT}^{(M)}$ elements. The probability that a sample will be verified is $p = 1 - FRR^{(M)}$, where $FRR^{(M)}$ is the False Rejection Rate for the master ensemble. The problem can be modeled using a binomial distribution. Therefore, on average:

$$k' = k\left(1 - FRR^{(M)}\right). \tag{1}$$

It is worth noting that the effectiveness of the master ensemble as supervisor towards the 'slave' matcher is related to its FRR value. If $\text{FRR}^{(\text{M})}=0$ ($k'=k$) the master ensemble is equivalent to a "true" supervisor because it will verify the class of all 'genuine' samples in D_u. If $\text{FRR}^{(\text{M})}>0$, then only $k'<k$ samples can be verified. Thus, only $k'<k$ samples will be added to the slave template set.

Only a finite number of samples are collectable in our physical system, so some samples appear indistinguishable. We are interested in samples that provide new information to the template set, so we can consider only distinguishable samples. Let d be the random variable representing the number of elements in k' that are different among themselves, and dn the random variable representing the number of elements in k' that are different and not present in the 'slave' template set. The pdf $p_d(d)$ can be modeled using a multinomial distribution:

$$p_d(d) = \binom{N_{TOT}^{(S)}}{d} \sum_{i=0}^{d} (-1)^i \binom{d}{i} \left(\frac{d-i}{N_{TOT}^{(S)}}\right)^{k'}. \tag{2}$$

The conditional pdf $p_{dn/d}(dn/d)$ – the probability to have dn different samples not present in the (S) template set if d samples are different among themselves – can be modeled using an hyper-geometric distribution:

$$p_{dn|d}(dn \mid d) = h\left(dn; N_{TOT}^{(S)}, (N_{TOT}^{(S)} - n), d\right). \tag{3}$$

where $N_{TOT}^{(S)}$ is the number of samples in the population, $(N_{TOT}^{(S)} - n)$ is the number of samples in the population that are classified as 'highly confident genuine samples' (i.e. samples actually belonging to the matched class), d is the number of items in the sample, dn is the number of items in the sample that are classified as 'highly confident genuine samples'. By using Eqs. (2) and (3), the pdf $p_{dn}(dn) = \sum_{d=1}^{k'} p_{dn/d}(dn/d)\, p_d(d)$ is derived. Here, we are interested to the expected value $E[dn]$. This represents the expected enlargement of the 'slave' template set due to the collection of k' elements. By assuming that for each co-update step the number $\Delta n^{(S)}$ of new and distinct samples added to the slave template set is about equal to its expected value $E[dn]$, since the correspondent variance var[dn] is very small:

$$\Delta n^{(S)} \approx E[dn] = \sum_{dn=0}^{k'} \left\{ dn \cdot p_{dn}(dn) \right\} = \left(1 - \left(1 - \frac{1}{N_{TOT}^{(S)}} \right)^{k\left(1-FRR^{(M)}\right)} \right) \left(N_{TOT}^{(S)} - n^{(S)} \right) . \qquad (4)$$

Let $n_0^{(S)}$ be the initial size of the (S) template set, and $n_{(i)}^{(S)}$ the size of the (S) template set at the *i*-th co-update step. The length of the (S) template set at the *i*-th step can be written as:

$$\begin{array}{l} n_{(0)}^{(S)} = n_0^{(S)} \\ n_{(i)}^{(S)} = n_{(i-1)}^{(S)} w + N_{TOT}^{(S)} (1 - w) \end{array} ; \quad w = \left(1 - 1/N_{TOT}^{(S)} \right)^{k\left(1-FRR^{(M)}\right)} . \qquad (5)$$

The weight *w* takes into account the effectiveness of the master template set as a supervisor. By recalling that in the ensemble of template matchers, as illustrated in Fig. 1, when a matcher plays the role of 'master' the other *c*-1 matchers play the role of 'slave', we can write the length of the j^{th} (S) template set at the *i*-th step

$$\begin{array}{l} n_{(0)}^{(S\,j)} = n_0^{(S\,j)} \\ n_{(i)}^{(S\,j)} = n_{(i-1)}^{(S\,j)} w^{(j)} + N_{TOT}^{(S\,j)} \left(1 - w^{(j)}\right) \end{array} ; \quad w^{(j)} = \left(1 - 1/N_{TOT}^{(S\,j)} \right)^{k\left(1-FRR^{(M)}\right)} . \qquad (6)$$

It is worth noting that while the effectiveness of the master template set as a supervisor depend on the $\text{FRR}^{(\text{M})}$ value, the final growth of each (S*j*) template set depend on each (S*j*) biometric itself, via the parameter $N_{TOT}^{(S\,j)}$. The $\text{FRR}^{(\text{M})}$ parameter is related to the behavior of the ensemble formed by the *c*-1 matchers.

The $\text{FRR}^{(\text{M})}$ value determine the system's dynamics, so it is worth to relate its value to the length of the master template set. In order to model the above value each j-th sample can be characterized by a $m_{ij} \geq 0$ connection degree with other samples, and the *i*-th class can be characterized by the sequence $\left\{ m_{i1}, m_{i2}, \ldots, m_{iN_{TOT}} \right\}$ of connection degrees. The "connection degree" is the number of examples that produce a similarity value over the updating threshold. There is a subset of samples for which this connection degree is zero. We refer to such samples with the term "isolated". We propose to characterize each client with two parameters: the fraction of samples that are isolated ($m_{ij} = 0$), namely, f_I, and the average connection degree $m_i = \left\langle \left\{ m_{ij} \right\} \right\rangle$ computed only for the 'connected' samples. Therefore, "difficult" classes can be modeled by a low value of m_i and high value of f_I, whilst high value of m_i and low value of f_I, are adopted for "easy" classes. Parameters m_i and f_I depend on the dataset, on the employed matchers, and on the ensemble formed by the 'master' matchers.

Isolated samples contribute to the FRR with a constant value equal to 1, because they cannot "match" to other samples. The contribute to connected sample is computed as follows. Let $\hat{x}$ be a connected sample with connection degree m_i. Let *r* be the number of the connected samples of $\hat{x}$ in the master template set, $0 \leq r \leq m_i$. The identity of $\hat{x}$ will be correctly verified if at least one among its m_i connected samples is in the master template set, that is, if $r>0$. If $r=0$ - events with probability p(r=0) - even connected samples cannot "match" with samples in the master template set, so their contribute to the FRR value will be $p(r=0)$.

When $n<(N_{TOT}-m_i)$ the random variable r will be modeled by the hypergeometric distribution $h\left(r;N_{TOT}^{(M)},m_i,n^{(M)}\right)$, and the FRR value depend on both isolated and connected samples. When the template set reaches the size $n=N_{TOT}-m_i$ at least one of the m_i connected samples will be in the master template set, so $p(r=0)=0$ and the FRR value depend only on the fraction of isolated samples f_I. In other words, there is a "saturation" value for FRR, that is FRR($n>N_{TOT}-m_i$)$=f_I$.

$$FRR^{(M)}=\begin{cases} f_I\cdot 1+f_C\cdot 0 & if\, n^{(M)}>N_{TOT}^{(M)}-m_i \\ f_I\cdot 1+f_C\cdot h(r=0;N_{TOT}^{(M)},m_i,n^{(M)}) & if\, n^{(M)}\le N_{TOT}^{(M)}-m_i \end{cases}. \tag{7}$$

Eq. (5-6) allows us to model the template set size increase during the co-update iterations, whilst eq. (7) models the FRR of the master ensembles. Notice that, in order to simplify the model, we adopted the average of m_{ij} for connected samples. This average value does not take into account isolated samples, but only connected samples. The above relationships can be used to predict the behaviour of co-update.

3 Experimental Results

In this Section, the case-study of a biometric verification system made up of a face and a fingerprint matchers is adopted for the experimental validation of the proposed model. In biometrics, a single sample is given by a couple of face and fingerprint images, from which two independent features sets are derived [2-3]. Images collected by the supervisor are the initial template set, which is called "gallery" in biometric applications.

Although the number of matchers is two, the model can be validated by hypothesizing that each matcher behaves as an ensemble of set of independent matchers applied to the same biometric. This agrees with observations in previous Section: the model proposed in [7] derives from the general framework proposed here. Therefore, the following conclusions can be applicable for the general case.

3.1 The Data Set

The data set adopted consists of 42 individuals composed of 20 face and fingerprint images for each individual, by keeping in mind the independence of face and fingerprint traits. The time span of both the collected data sets spans over one year. Forty-two frontal face images with 20 instances representing significant illuminations changes and variations in facial expressions per person were used from the Equinox corporation database [11]. The fingerprint data set has been collected by the authors using Biometrika Fx2000 optical sensor. The images are acquired with variations in pressure, moisture and time interval to represent large intra-class variations. The results are computed on five random coupling of face and fingerprint datasets and are averaged. Whilst minutiae are simply extracted from the fingerprint images after commonly used processing algorithms [2], PCA is computed on the whole data set and applied to face images in order to reduce the size of the overall feature space. 95% of energy is retained according to the current literature [3].

It is worth noting that fingerprint and face data sets are strongly different in terms of environmental conditions: the face one is notably "simpler" than the fingerprint one. We adopted so different data sets in order to show the effect of intra-class variations on the model prediction ability.

3.2 Experimental Results

First of all, we implemented a simple bi-modal identification system made up of a PCA-based face matcher and a fingerprint matcher using the "String" matching algorithm ("String" is based on minutiae points). We used the standard versions of these two recognition algorithms [2-3]. Then, we implemented the template co-update algorithm in Figure 1. Both the eigenspace of the PCA-based face matcher and the co-update threshold value at FAR = 0% are computed using the whole dataset, and has not been updated during the co-update process. This approach is common to some template update practices [5-9]. The update of face and fingerprint templates is performed simply by adding new examples to the user's template set.

In the experiments the initial template set has been set as follows. We selected, as the initial template in D_l for the i-th client, a sample whose connection degree is exactly m_i. In other word, the initial template is a connected sample 'near' to other m_i samples. In the proposed model the i-th client is characterized by (f_C, m_i) so this initial template is representative for the client. The rationale for this choice is to exclude outliers from the initial template set, likewise to what happens in real situations, where the initial template is chosen in completely supervised fashion.

We simulated the acquisition of a batch set D_u by generating several sets of k=10 couples – face and fingerprint – of 'genuine' examples, drawn with replacement from a homogeneous population of $N_{TOT}^{(face)} = N_{TOT}^{(fingerprint)}$ = 20 samples. We are aware that adopted database size may not be very well appropriate for the task, but it respects, on average, the size adopted in other template update works reported in literature [5-9].

In order to set the correct parameters in the proposed model, for each client and for each biometric we computed a) the value f_c, that is, the fraction of samples that produce a 'score' over the threshold S^*, and b) the value m_i, that is, the integer nearest to the average connection degree of the connected samples. For each client, results are averaged on ten trials (experimental values) and predicted using parameters evaluated on the related training set. Table 1 gives a summary of these parameters. In particular, it is shown the ratio between the average fraction of connected samples and the related standard deviation (second column). The ratio between the average connection degree and the related standard deviation is also shown.

Table 1. Ratio between the average fraction of connected samples (f_c) and related standard deviation. The same ratio related to the connection degree m_s is also shown for face and fingerprint systems.

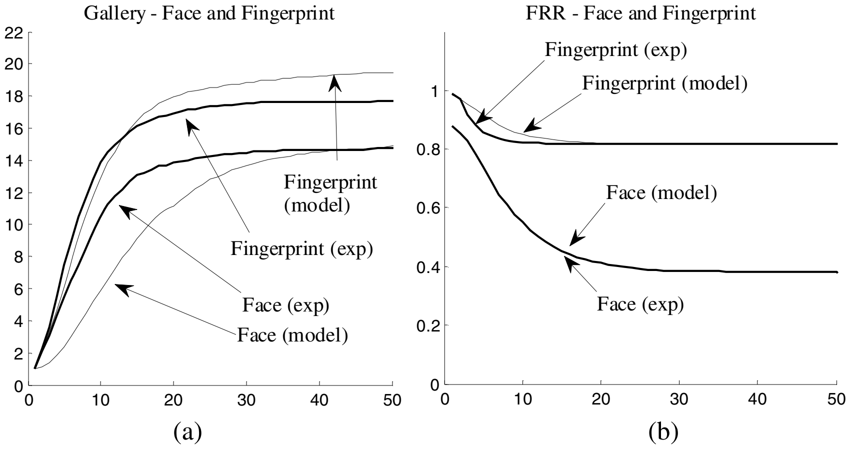
	E[f_c]/devst[f_c]	**E[m]/devst[m]**
Face	0,62	3,15
Fingerprint	0,71	1,33

It is worth noting that the difference among considered biometric systems is not related to the fraction of connected samples. In fact, the value of fingerprint and face matchers is about the same. This mean that, with respect to the average value, the related standard deviation is such that the differences two systems is not appreciable. This is confirmed by the average value of f_c, which is 0,62 for the face matcher and 0,71 for the fingerprint one.

Strong differences can be appreciated by observing the ratio between the average connection degree (*m*) and the related standard deviation. This indicated that face matcher exhibit a connection degree higher than that of fingerprint one, and a lower standard deviation. Therefore, face matcher can be considered (maybe due to Equinox data) "easier" than fingerprint one, thus it is expected a significant performance improvement for the face matcher, by using the co-update approach..

This hypothesis is confirmed by Figs. 2, which compares the values obtained on test data with the values predicted by the model. Reported trend is obtained by averaging values over all the clients. For each iteration of the algorithm depicted in Figs. 2, the FRR value evaluated at zeroFAR and the template set sizes are reported. The model is able to predict the experimental performance, with a negligible difference. For example, in a real scenario, due to the presence of isolated samples, the true value of FRR will be greater than 0 (FRR=1 if all the samples in template set are isolated), and this is respected by proposed model.

Fig. 2 shows the average values over all the client. Discrepancy between predicted and experimental values can be justified recalling that the proposed model predicts the system's performance for each client separately. Moreover, the whole connection graph that characterizes each client is modelled using only two parameters (m_i and f_I). The saturation of theoretical and experimental curves is driven by eqs. (6-7). In particular, FRR of the master matcher is responsible of the weight *w*, which tends to decrease as the template set size of both matchers increases. Thus, the size of galleries



(a) (b)

Fig. 2. Experimental trend and predicted values of template set sizes (a) and FRR (b) for fingerprint and face biometrics. Reported trend is obtained averaging values over all the clients.

must converge to N_{TOT} if m_i is not zero. At the end of co-update process, FRR(n=N_{TOT})=f_I. It is worth noting that the correct prediction of FRR values is matter of primary importance both for the validation of the model and for the designers, that need relevant information about the performance of the system. With regard to this issue, the FRR predicted by the proposed framework does not depend on the number of matchers, but only the template set size and related parameters (connection degree and fraction of connected samples).

An important problem to solve is how to compute the above parameters. At the state of our knowledge, this can be only done by collecting an additional set of examples. But this implies to verify that these parameters are not dependent on intra-class variations, but to intrinsic characteristic of each client. This will be matter of a future work.

4 Conclusions

In this paper we showed that the theoretical framework of template co-update can be derived from a general model proposed here. Preliminary experiments on a biometrics-based case-study have shown the effectiveness of the model. A further validation step is necessary by increasing the number of biometrics.

Other issues still remain, but we believe that the template co-update algorithm is well explained by proposed model. Therefore, it is suitable for designing classification systems, based on template matching, which can improve with use.

References

1. Duda, R., Hart, P., Stork, D.: Pattern Classification. Wiley-Interscience, Hoboken (2000)
2. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of fingerprint recognition. Springer, Heidelberg (2003)
3. Li, S.Z., Jain, A.K. (eds.): Handbook of face recognition. Springer, Heidelberg (2005)
4. Ross, A., Nandakumar, K., Jain, A.K.: Handbook of Multibiometrics. Springer, Heidelberg (2006)
5. Zhu, X.: Semi-supervised learning literature survey, Technical report, Computer Sciences TR 1530, Univ. Wisconsis, Madison, USA (2006)
6. Roli, F., Marcialis, G.L.: Semi-supervised PCA-based face recognition using self-training. In: Yeung, D.-Y., Kwok, J.T., Fred, A., Roli, F., de Ridder, D. (eds.) SSPR 2006 and SPR 2006. LNCS, vol. 4109, pp. 560–568. Springer, Heidelberg (2006)
7. Didaci, L., Marcialis, G.L., Roli, F.: Modelling FRR of Biometric Verification Systems using the Template Co-update Algorithm. In: 3rd IAPR/IEEE Int. Conference on Biometrics ICB 2009, Alghero (Italy), June, 2-5. LNCS, Springer, Heidelberg (2009) (in press)
8. Roli, F., Didaci, L., Marcialis, G.L.: Adaptive biometric systems that can improve with use. In: Ratha, N., Govindaraju, V. (eds.) Advances in Biometrics: Sensors, Systems and Algorithms, pp. 447–471. Springer, Heidelberg (2008)
9. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Proc. of the Workshop on Computational Learning Theory, pp. 92–100 (1998)
10. Kittler, J., Hatef, M., Duin, R., Matas, J.: On Combining Classifiers. IEEE Trans. On PAMI 20(3), 226–239 (1998)
11. `http://www.equinoxsensors.com/products/HID.html`