Analysis of multimodal binary detection systems based on dependent/independent modalities

Oleksiy Koval
CUI-University of Geneva,
24, rue General Dufour,
1211 Geneva, Switzerland,
Oleksiy.Koval@cui.unige.ch

Sviatoslav Voloshynovskiy
CUI-University of Geneva,
24, rue General Dufour,
1211 Geneva, Switzerland,
svolos@cui.unige.ch

Thierry Pun
CUI-University of Geneva,
24, rue General Dufour,
1211 Geneva, Switzerland,
Thierry.Pun@cui.unige.ch

***Abstract*— Performance limits of multimodal detection systems are analyzed in this paper. Two main setups are considered, i.e., based on fusion of dependent and independent modalities, respectively. The analysis is performed in terms of attainable probability of detection errors characterized by the corresponding error exponents. It is demonstrated that an expected performance gain from fusion of dependent modalities is superior than in the case when one fuses independent signals. In order to quantify the efficiency of dependent modality fusion versus the independent case, the problem analysis is performed in the Gaussian formulation.**

I. Introduction

Hypothesis testing is widely used for modeling, design and performance optimization of of various practical systems that include but not limited to radar imaginary, speaker detection, person identification, etc. The main goal of the test consists in defining, which of possible alternatives takes place based on the observed data. In the very simplest case of binary hypothesis testing that will be considered in this paper, there are only two alternatives to be discriminated.

Intuitively, exploitation of the complete data set of multimodal signals should lead to more accurate detection performance than in the case of incomplete data set exploitation. However, in some practial situations like biometric person identification till recent past only part of the available biometric data was exploited in order to reduce the identification system complexity and overall cost [1].

However, such a simplification to monomodal biometric identification systems leads to a performance that does not satisfy sophisticated security requirements of the modern society. That is why the main trends of current biometric person identification is to deviate toward multibiometric design where joint exploitation of various biometric data (like facial photos and fingerprints) should lead to performance enhancement versus unimodal setup [1], [2], [3].

Fusion of multibiometrics can be performed on various structural levels of an identification system [4]: sensor level, feature level, match score level, rank level and decision level. Due to the data processing inequality [5], expected performance improvement will be the highest if fusion is performed on the sensor level or on the level of the observed data. Despite this approach has not received significant attention in practice due to various technological aspects [6], its information-theoretic analysis that remains an open problem should provide an attainable upper bound on such a system performance. The second motivating aspect of the current research consists in the uncertainty about the influence of the dependence/independence of the fused multimodal signals used in the detection framework on the attained theoretical system performance. The existing results in this direction concern the correlation structure of these data. It might be unexpectable, but there does not exist a unique treatment of this problem. According to the results presented in [7], mutimodal detection based on correlated data does not always lead to the overall performance improvement versus combining independent signals. Oppositely, as it is demonstrated in [8], taking into account correlation between the multimodal signals one can expect some performance improvement.

Therefore, motivated by the existing gap in the theoretical performance analysis of a binary multimodal detection that is considered as a theoretical model of real multimodal biometric person identification systems, we would like to formulate the main goal of this paper in the justification of its performance in terms of error exponents for both cases of dependent and independent multimodal cases.

The rest of the paper has the following structure. We present the problem formulation of the theoretical analysis of binary multimodal detection in Section 2. Performance analysis of multimodal detection based on independent and dependent signals is performed in Section 3. Conclusion and future research perspectives are formulated in Section 4.

Notations We use capital letters to denote scalar random variables X and corresponding small letters x to denote their realizations. The superscript N is used to designate length-N vectors $x^N = [x[1], x[2], ..., x[N]]$ with k^{th} element $x[k]$. We use $X \sim p_X(x)$ or simply $X \sim p(x)$ to indicate that a random variable X is distributed according to $p_X(x)$. The mathematical expectation of a random variable $X \sim p_X(x)$ is denoted by μ_X and σ_X^2 denotes the variance of X. We use Σ to denote a covariance matrix. Correlation coefficient between two random variables is designated by ρ. Calligraphic fonts $\mathcal{X}$ denote sets $X \in \mathcal{X}$ and $|\mathcal{X}|$ denotes the cardinality of set $\mathcal{X}$. Superscript T stays for matrix transposition.

II. PROBLEM FORMULATION

We model a problem of binary multimodal detection as a binary hypothesis testing (Fig. 1). It is supposed that the source of multimodal signals modeled by a joint probability distribution $p(x^N, y^N)$ generates a pair of N/length vectors $X^N \in \mathcal{X}^N$ and $Y^N \in \mathcal{Y}^N$, $(X^N, Y^N) \sim q(x^N, y^N)$. It should be noted that in the general case, the length of the observed data vectors is not necessary the same (Fig. 2) and this particular selection was made for analysis simplicity sake. The task of the hypothesis testing block that observes this pair of vectors is to perform a test η in order to decide which of two alternative cases is represented by the mentioned pair.Thus, a binary multimodal detection system consists of the set $\{X^N, Y^N\}$ and a hypothesis test:

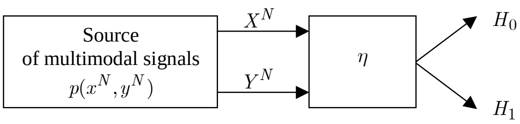

Fig. 1. Binary hypothesis testing system.

$$\eta : \mathcal{X}^N \times \mathcal{Y}^N \to \{0,1\}, \qquad (1)$$

where {0,1} designates the alternative hypotheses that might take place, H_0 and H_1, respectively. Thus, the main task of multimodal binary hypothesis testing is to decide which one of the two hypotheses is true given (X^N, Y^N). This task is performed according to the following test:

$$\begin{cases} H_0, (X^N, Y^N) \sim p^0(x^N, y^N), \\ H_1, (X^N, Y^N) \sim p^1(x^N, y^N), \end{cases} \qquad (2)$$

where $H_0 \sim p^0(x^N, y^N) = \prod_{i=1}^N p^0(x_i, y_i), H_1 \sim p^1(x^N, y^N) = \prod_{i=1} p^1(x_i, y_i)$ denote a priori statistical models on alternative hypotheses.

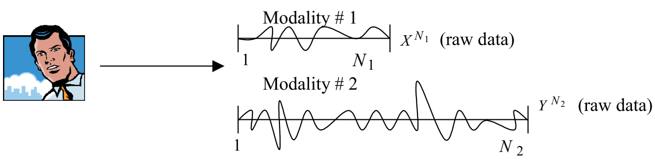

Fig. 2. Multimodal observations: the vectors of different lengths might be observed.

In order to attain the lowest probabilities of detection error, it is assumed that the test follows the Neyman-Pearson rule [5]. These errors are of two kinds, type I error or a false alarm, denoted as P_f, occurs if the hypothesis test decides that H_1 is true while H_0 is in force, and type II error or a miss, denoted as P_m, that occurs if an uncorrect decision about H_0 is made. These error probabilities are defined as follows:

$$P_m = \Pr[\eta \leq T | H_1], \qquad (3)$$

$$P_f = \Pr[\eta \geq T | H_0]. \qquad (4)$$

According to the Neyman-Pearson lemma, for a given maximal tolerable probability P_f, P_m can be minimized iff the log-likelihood ratio test is applied. This test is defined as:

$$\eta = N\left\{D(q(x,y) \,||\, p^1(x,y)) - D(q(x,y) \,||\, p^0(x,y))\right\}$$
$$\geq \log_2 T. \qquad (5)$$

Here it is assumed that $p(x^N, y^N) = \prod_{i=1}^N p(x_i, y_i)$, and thus the empirical distribution and a priory hypothesis models are defined as $q(x^N, y^N) = \prod_{i=1}^N q(x_i, y_i)$, $p^0(x^N, y^n) = \prod_{i=1}^N p^0(x_i, y_i)$, $p^1(x^N, y^n) = \prod_{i=1}^N p^1(x_i, y_i)$; $D(.\,||\,.)$ stays for a relative entropy between two distributions [5] and T designates a predefined threshold.

In case error probabilities of false alarm and miss are defined according to (3) and (4), one has [9]:

$$P_m \log \frac{P_m}{1 - P_f} + (1 - P_m) \log \frac{1 - P_m}{P_f}$$
$$\leq D(p^1(x^N, y^N) \,||\, p^0(x^N, y^N)) \qquad (6)$$

If one fixes the probability of $P_m = 0$ in (6), a lower bound on P_f that increases with decrease of relative entropy $D(p^1(x^N, y^N) \,||\, p^0(x^N, y^N))$ is obtained:

$$P_m \leq 2^{-D(p^1(x^N, y^N)||p^0(x^N, y^N)))}. \qquad (7)$$

Therefore, it is evident that in order to optimize the performance of a multimodal binary hypothesis testing, one should maximize $D(p^1(x^N, y^N) \,||\, p^0(x^N, y^N))$. In order to achieve optimality in terms of the Bayessian probability of error, $P_e = \pi_I P_f + \pi_{II} P_m$, where π_I, π_{II} stay for costs of making the error of type I and II, respectively, the so-called J-divergence, $J = D(p^1(x^N, y^N) \,||\, p^0(x^N, y^N)) + D(p^0(x^N, y^N) \,||\, p^1(x^N, y^N))$ should be maximized [10].

Finally, the complete system performance analysis can be performed based on Stein lemma [5]. According to this lemma the performance of the Neyman-Pearson test is defined as:

$$P_f \sim 2^{-N\left[D(p^1(x,y))||D(p^0(x,y))\right]}, \text{for a fixed } P_m, \qquad (8)$$

$$P_m \sim 2^{-N\left[D(p^0(x,y))||D(p^1(x,y))\right]}, \text{for a fixed } P_f. \qquad (9)$$

Thus, the overall multimodal binary detection system performance is determined by the corresponding relative entropies defined with respect to the prior distributions on alternative hypotheses.

The following sections contain the analisis of the modality dependence impact on the corresponding probabilities of error.

III. PERFORMANCE ANALYSIS

A. Independent modalities

Here we assume that $p^1(x,y) = p^1(x)p^1(y); p^0(x,y) = p^0(x)p^0(y)$. The chain rule for relative entropies leads to:

$$D(p^1(x,y)) \,||\, D(p^0(x,y))$$
$$= D(p^1(y) \,||\, p^0(y)) + D(p^1(x) \,||\, p^0(x)), \quad (10)$$

$$D(p^0(x,y)) \,||\, D(p^1(x,y))$$
$$= D(p^0(y) \,||\, p^1(y)) + D(p^0(x) \,||\, p^1(y)). \quad (11)$$

The corresponding bounds on the probabilities of error are:

$$P_f \sim 2^{-N\left[D(p^1(y)||p^0(y))+D(p^1(x)||p^0(x))\right]}, \quad (12)$$

for a fixed and arbitrary small P_m,

$$P_m \sim 2^{-N\left[D(p^0(y)||p^1(y))+D(p^0(x)||p^1(y))\right]}, \quad (13)$$

for a fixed and arbitrary small P_f. Thus, one can conclude that performance probabilities of incorrect detection measured in terms of error exponents decrease with a number of exploited multimodal signals.

B. Dependent modalities

In the case of dependent modalities ($p(x,y) \neq p(x)p(y)$) the bounds on the probabilities of error are determined by (8) and (9).

Applying the chain rule for the relative entropy one obtains:

$$D(p^1(x,y) \,||\, p^0(x,y))$$
$$= D(p^1(y) \,||\, p^0(y)) + D(p^1(x\,|y\,) \,||\, p^0(y\,|x\,)), \quad (14)$$

$$D(p^0(x,y) \,||\, p^1(x,y))$$
$$= D(p^0(y) \,||\, p^1(y)) + D(p^0(x\,|y\,) \,||\, p^1(x\,|y\,)). \quad (15)$$

Comparing performance bounds for dependent (8) and (9) and independent (14) and (15), one needs to consider the following quantities:

$$D(p^0(x) \,||\, p^1(x)) \text{ vs. } D(p^0(x\,|y\,) \,||\, p^1(x\,|y\,)), \quad (16)$$

$$D(p^1(x) \,||\, p^0(x)) \text{ vs. } D(p^1(x\,|y\,) \,||\, p^0(x\,|y\,)). \quad (17)$$

If, $D(p^0(x) \,||\, p^1(x)) \leq D(p^0(y\,|x\,) \,||\, p^1(x\,|y\,))$, and $D(p^1(x) \,||\, p^0(x)) \leq D(p^1(x\,|y\,) \,||\, p^0(x\,|y\,))$, one can conclude that systems exploiting statistically dependent signals a better performance than ones working with independent signals.

Lemma: Conditioning does not reduce relative entropy.

Proof.

$$\begin{aligned} & D(p^0(y\,|x\,) \,||\, p^1(y\,|x\,)) - D(p^0(x) \,||\, p^1(x)) \\ &= \textstyle\sum_x \sum_y p^1(x,y) \log \frac{p^1(x|y\,)}{p^0(x|y\,)} - \sum_x p^1(x) \log \frac{p^1(x)}{p^0_X(x)} \\ &= \textstyle\sum_x \sum_y p^1(x,y) \log \frac{p^1(x|y\,)}{p^0(x|y\,)} - \sum_x \sum_y p^1(x,y) \log \frac{p^1(x)}{p^0(x)} \\ &= \textstyle\sum_x \sum_y p^1(x,y) \log \frac{p^1(x|y\,)p^0(x)}{p^0(x|y\,)p^1(x)} \\ &\geq 1 - \textstyle\sum_x \frac{p^1(x)}{p^0(x)} \sum_y p^1(y)p^0(x\,|y\,) \\ &= 1 - \textstyle\sum_x \frac{p^1(x)}{p^0(x)} p^0(x) = 0, \end{aligned} \quad (18)$$

where the only inequality in (7) is due to $\log(x) \geq 1 - \frac{1}{x}$.

Thus, based on (6) one can state that multimodal detection systems based on fusion of independent signals have higher theoretically attainable probabilities f detection errors bounded by corresponding error exponents than one expects in the case of dependent signals.

C. Bivariate Gaussian case.

In order to quantify the expected performance gain one can expect from fusion of dependent modalities in binary hypothesis testing, it was assumed that the priors on alternative hypotheses follow bivariate Gaussian distributions:

$$p^1(x,y) = \frac{1}{2\pi\sqrt{\det(\Sigma_1)}} \exp\Big\{ -\frac{1}{2} \left[x-\mu_{X_1}, y-\mu_{Y_1}\right]^T \times \Sigma_1^{-1} \left[x-\mu_{X_1}, y-\mu_{Y_1}\right] \Big\}; \quad (19)$$

$$p^0(x,y) = \frac{1}{2\pi\sqrt{\det(\Sigma_0)}} \exp\Big\{ -\frac{1}{2} \left[x-\mu_{X_0}, y-\mu_{Y_0}\right]^T \times \Sigma_0^{-1} \left[x-\mu_{X_0}, y-\mu_{Y_0}\right] \Big\}, \quad (20)$$

with mean vectors $[\mu_{X_1}, \mu_{Y_1}]$, $[\mu_{X_0}, \mu_{Y_0}]$ and covariance matrices

$$\Sigma_1 = \begin{pmatrix} \sigma^2_{X_1} & \rho\sigma_{X_1}\sigma_{Y_1} \\ \rho\sigma_{X_1}\sigma_{Y_1} & \sigma^2_{Y_1} \end{pmatrix};$$

$$\Sigma_0 = \begin{pmatrix} \sigma^2_{X_0} & \rho\sigma_{X_0}\sigma_{Y_0} \\ \rho\sigma_{X_0}\sigma_{Y_0} & \sigma^2_{Y_0} \end{pmatrix},$$

where ρ is a correlation coefficient.

The joint relative entropies that define corresponding probabilities of detection error are defined as:

$$P_f : D(p^1(x,y) \,||\, p^0(x,y)) =$$
$$\frac{1}{2}\Big\{ \log_2 \frac{\det(\Sigma_1)}{\det(\Sigma_0)} + tr\left[\Sigma_1^{-1}\Sigma_0\right] + \left[\mu_{X_0}-\mu_{X_1}, \mu_{Y_0}-\mu_{Y_1}\right] \times \Sigma_0^{-1} \left[\mu_{X_0}-\mu_{X_1}, \mu_{Y_0}-\mu_{Y_1}\right]^T \Big\}; \quad (21)$$

$$P_m : D(p^0(x,y) \,||\, p^1(x,y)) =$$
$$\frac{1}{2}\Big\{ \log_2 \frac{\det(\Sigma_0)}{\det(\Sigma_1)} + tr\left[\Sigma_0^{-1}\Sigma_1\right] + \left[\mu_{X_1}-\mu_{X_0}, \mu_{Y_1}-\mu_{Y_0}\right] \times \Sigma_1^{-1} \left[\mu_{X_1}-\mu_{X_0}, \mu_{Y_1}-\mu_{Y_0}\right]^T \Big\}, \quad (22)$$

where inverse covariance matrices Σ_1^{-1} and Σ_0^{-1} defined in the following way:

$$\Sigma_1^{-1} = \frac{1}{\sigma_{X_1}^2 \sigma_{Y_1}^2 (1-\rho^2)} \begin{pmatrix} \sigma_{Y_1}^2 & -\rho\sigma_{X_1}\sigma_{Y_1} \\ -\rho\sigma_{X_1}\sigma_{Y_1} & \sigma_{X_1}^2 \end{pmatrix},$$

$$\Sigma_0^{-1} = \frac{1}{\sigma_{X_0}^2 \sigma_{Y_0}^2 (1-\rho^2)} \begin{pmatrix} \sigma_{Y_0}^2 & -\rho\sigma_{X_0}\sigma_{Y_0} \\ -\rho\sigma_{X_0}\sigma_{Y_0} & \sigma_{X_0}^2 \end{pmatrix}.$$

In order to demonstrate possible performance gain, the parameters of a priory distributions $p^1(x,y)$ and $p^0(x,y)$ where selected to be $\mu_{X_0} = 10, \mu_{X_1} = 20, \sigma_{X_0}^2 = 36, \sigma_{X_1}^2 = 16, \mu_{Y_0} = 4, \mu_{Y_1} = 8, \sigma_{Y_0}^2 = 4, \sigma_{Y_1}^2 = 6$. The behavior of $D(p^1(x,y)\,||\,p^0(x,y))$ and $D(p^0(x,y)\,||\,p^1(x,y))$ as functions of the correlation coefficient ρ was analyzed. The obtained results are shown in Figure 3. They entirely confirm our theoretical findings.

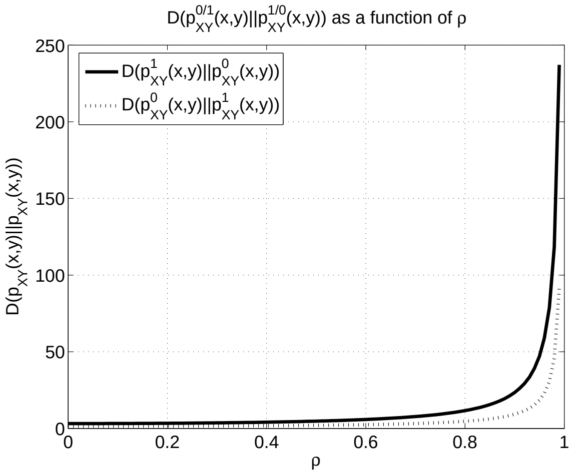

Fig. 3. $D(p^1(x,y)\,||\,p^0(x,y))$ and $D(p^0(x,y)\,||\,p^1(x,y))$ as functions of correlation coefficient ρ.

IV. Conclusions and future research perspectives

In this paper we considered the problem of performance analysis of binary multimodal detection. In particular, we considered two particular problem formulations where hypothesis test is performed based on independent and dependent multimodal signals. Corresponding lower bounds on the probabilities of miss and false alarm in terms of error exponents for both setups were developed and it was theoretically proved that dependence between multimodal signals leads to the overall enhanced system performance versus the setup with independent modalities. For demonstration purpose, the bivariate Gaussian problem formulation was analyzed and respective results quantifying performance improvement were obtained. Since in the case of Gaussian data independence is equivalent to the uncorrelation, one can conclude that fusion of correlated modalities leads to a higher accuracy in classification problem. In particular, relative entropies that define the corresponding probability of errors, i.e., probability of false alarm and probability of miss, are non-decreasing monotonic functions of the correlation coefficient ρ on the interval [0,1].

As a possible direction for the obtained results extension we see its application to a general multimodal detection system architecture that assumes multiple hypothesis formulation analysis (Fig. 4).

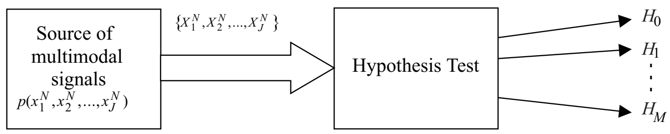

Fig. 4. Multiple hypothesis testing system.

Another potential future research line consists in the practical validation of the developed framework in multimodal person identification application using ID cards that contain embedded biometric data and personal data. Our goal is to develop a general system structure and to evaluate its performance.

V. Acknowledgements

This paper was partially supported by SNF Professeur Boursier grant PP002–68653, by the European Commission through the IST Programme under contract IST-2002-507932-ECRYPT and European Commission through sixth framework program under the number FP6-507609 (SIMILAR) and Swiss IM2 projects. The authors are thankful to the members of SIP group, University of Geneva for many stimulating and interesting discussions.

References

[1] J. L. Wayman, A. K. Jain, D. Maltoni, and D. Maio, *Biometric Systems: Technology, Design and Performance Evaluation*, Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2004.

[2] L. Hong, A. K. Jain, and S. Pankanti, "Can multibiometrics improve performance?," Tech. Rep. MSU-CSE-99-39, Department of Computer Science, Michigan State University, East Lansing, Michigan, December 1999.

[3] L. Hong and A. K. Jain, "Integrating faces and fingerprints for personal identification," in *ACCV '98: Proceedings of the Third Asian Conference on Computer Vision-Volume I*, pp. 16–23, Springer-Verlag, (London, UK), 1997.

[4] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: A tool for information security," *IEEE Trans. on Information Forensics and Security* **1**, pp. 125–143, June 2006.

[5] T. Cover and J. Thomas, *Elements of Information Theory.*, Wiley and Sons, New York, 1991.

[6] A. Ross, R. G. Hong, A. K. Jain, and S. Pankanti, "Feature level fusion using hand and face biometrics," in *SPIE Conf. Biometric Technology for Human Identification II*, pp. 196–204, 2005.

[7] N. Poh and S. Bengio, "ow do correlation and variance of base classifiers affect fusion in biometric authentication tasks?," *IEEE Transactions on Signal Processing* **53**, pp. 4384–4396, November 2005.

[8] O. Ushmaev and S. Novikov, "Biometric fusion: Robust approach," in *Workshop on Multimodal User. Authentication (MMUA 2006)*, (Toulouse, France), 2006.

[9] R. E. Blahut, *Principles and practice of information theory*, Addison Wesley Publ. Co., 1987.

[10] H. V. Poor, *An Introduction to Detection and Estimation Theory*, Springer-Verlag, 1994.